

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Genesed_36_Q03665	+	44.00	113.92	95.86	2455	Sequence homologous to Drospophila
N_Genesed_36_X20561	+	44.00	108.58	190.22	4354	Polynucleotide sequence from the
N_Genesed_36_Q20765	+	44.00	105.47	283.39	6062	PAD-CMV19. O-glycosylated alpha
N_Genesed_36_Q20732	+	44.00	105.47	283.39	6062	PAD-CMV19 expression vector. C
N_Genesed_36_Q02683	+	44.00	104.94	303.33	6414	PAD-CMV1 contig. A Tumour Neuro
N_Genesed_36_Q02766	+	44.00	104.94	303.33	6414	PAD-CMV1. O-glycosylated alpha
N_Genesed_36_Q20733	+	44.00	104.94	303.33	6414	PAD-CMV1 expression vector. O-
N_Genesed_36_Q20733	+	44.00	104.94	303.33	6414	PAD-CMV1 expression vector. O-
N_Genesed_36_Q43361	+	44.00	104.63	315.67	6630	Sequence of circular plasmid P
N_Genesed_36_Q20551	+	44.00	104.13	336.60	6993	Polynucleotide sequence from
N_Genesed_36_Q34545	+	43.00	124.25	25.49	589	-4aa IL-6 SSSS mutin. New cyst
N_Genesed_36_Q89359	+	43.00	124.25	25.49	589	Human interleukin-6 (-Ala1), Se
N_Genesed_36_T39892	+	43.00	124.25	25.49	589	Coding sequence for -4aa SSCC m
N_Genesed_36_V37110	-	43.00	126.18	39.42	846	Antibiotic resistance gene blaZ
N_Genesed_36_T07329	+	43.00	116.85	71.76	1391	Human calcium signal-modulat
N_Genesed_36_X13498	+	43.00	114.81	85.58	1610	Enterococcus faecalis genome c
N_Genesed_36_X40398	+	43.00	108.43	193.98	3176	LeckShs domain-combining prote
N_Genesed_36_Q88185	+	43.00	103.71	355.20	5248	Bacillus circulans CTPase comp
N_Genesed_36_T93353	+	42.00	124.45	24.86	414	EST locus z82b09.s1 from CDNA
N_Genesed_36_T93353	+	42.00	124.11	25.94	429	EST locus z84a07.s1 from CDNA
N_Genesed_36_T93342	+	42.00	123.94	26.53	437	EST locus YV23e03.s1 from CDNA
N_Genesed_36_T93352	+	42.00	123.60	27.70	453	EST locus z84a07.r1 from CDNA
N_Genesed_36_T93344	+	42.00	123.18	29.26	474	EST locus yfile03.s1 from CDNA
N_Genesed_36_X14180	-	42.00	113.28	104.14	1360	H. pylori GHP0 718 gene. New
N_Genesed_36_T18796	-	42.00	111.64	128.47	1619	Human immunophilin clone hm51
N_Genesed_36_V45137	-	42.00	111.04	138.67	1725	Homo sapiens P-TEN tumour supp
N_Genesed_36_T93354	+	42.00	110.79	143.33	1773	Gene of IMAGE clone 264611. Nu
N_Genesed_36_Q07339	-	42.00	110.73	144.40	1784	Dual specificity phosphatase F
N_Genesed_36_Q13865	+	42.00	109.27	174.13	2084	PRP 140 gene and regulatory re
N_Genesed_36_Q14235	-	42.00	107.74	211.70	2451	A. faecalis penicillin acylase
N_Genesed_36_V46392	+	42.00	105.36	287.50	3160	Human tumour suppressor TS10q2
N_Genesed_36_Q13864	+	42.00	104.33	328.07	3526	PRP 378 gene and regulatory re
N_Genesed_36_V174765	+	42.00	102.21	430.47	4418	Staphylococcus aureus contig S
N_Genesed_36_T09311	+	42.00	98.43	699.50	6611	Mycobacteriophage genomic DS6A
N_Genesed_36_T51223	+	42.00	98.43	699.50	6611	NheI-D fragment of DS6A genome
N_Genesed_36_T66129	+	42.00	98.43	699.50	6611	Mycobacteriophage DS6A NheI fr
N_Genesed_36_T70491	+	42.00	98.43	699.50	6611	Mycobacteriophage DS6A NheI-D
N_Genesed_36_T37535	+	41.00	113.00	107.87	1005	Human secreted protein cDNA fr
N_Genesed_36_T71711	-	41.00	105.65	276.86	2198	Cellulomonas fimi endoglucanase
N_Genesed_36_X13111	-	41.00	101.29	484.20	3496	Enterococcus faecalis genome c
N_Genesed_36_X10416	-	41.00	95.23	1.1e+03	6656	Mutant protease gene (delta137

```

FT      /label= DHFR
FT      /note= "intron 1"
FT      2151..2168
FT      /tag= m
FT      /note= "binding site for EBI-1857 (Q22522)"
FT      2344..2821
FT      /tag= n
FT      /label= DHFR

```

```

1 3 001
FT 2022. 03474
FT /*tag= o
FT /label= DHFR
FT 3475. 3812
FT rep_origin
FT /*tag= p

```

```

FT /label= 3440
FT misc_feature 3813. .6055
FT /*tag= q
FT /note= "bluescript portion"
FT misc_feature 3813. .4291

```

chr	start	end	strand	cds	note	region
1	1000000	1000000	+		"M13 intergenic region (M13ori)"	
FT						
FT						
FT						

```
FI      misc feature
FT      6038.  .6062
        /label= beta-lactamase
```

DE4021917-A.  
PN  
FT  
PT  
L  
/note= "binding site for EBI-2134 (Q22518)"  
/tag-

10-JAN-1992.  
FU  
PF 10-JUL-1990; 021917.  
PR 10-JUL-1990; DE-021917.  
PA (BOEH ) BOEHRINGER INGELHEI.  
PI Himmeler A, Adolf G;  
WPI: 92-035485/04.  
DR  
PT Q-glycosylated alpha-interferon, used as medicament - isolated

PT contg. a suitable expression plasmid  
PT following secretion into conditioned  
ps medium or mammalian cells  
ps disclosure: Fig 7(A-D), 24pp. German

CC pAD-CMV9 includes CMV promoter  
CC acceptor sites, SV40 early and  
CC cloning site. The 5' non-coding

```
1. .590
/*tag= c
/not= "any enhance and protect our
```

```
722. .740
/*tag= d
```

```
741. .805
/*tag= e
```

```
836. .853
/*tag= f
```

```
862. .922
/*tag= g
```

```
/taget=mult cloning_size  
923. .1055  
/*tag= h
```

```

/notice= FROM SV40
1056. .1953
/*tag= i

```

1056. 1953  
hamster DHFR gene"  
/note= promoter and 3' non-coding

```

/*tag=
/note= "promoter and 5' non-coding
hamster buff gene"

```

```
1954. .2039
/*tag= k
label= DUFF
```

```

/note= "exon 1"
2040. .2333

```

Page 2

FT	/label= DHFR
FT	/note= "intron 1"
FT	2151. .2168
FT	misc_feature
FT	/*tag= m
FT	/note= "binding si
FT	2344. .2821
FT	exon
FT	/*tag= n
FT	/label= DHFR
FT	/note= "exons 2-6"
FT	2822. .3474
FT	3'utr
FT	/*tag= O
FT	/label= DHFR
FT	3475. .3812
FT	rep_origin
FT	/*tag= p
FT	/label= SV40
FT	3813. .6055
FT	misc_feature
FT	/*tag= q
FT	/note= "paluescrip
FT	3813. .4291
FT	misc_feature
FT	/*tag= r
FT	/note= "M13 intergr
FT	4423. .5283
FT	cds
FT	/*tag= s
FT	/label= beta-lacta
FT	6038. .6082
FT	misc_feature
FT	/*tag= t
FT	/note= "binding si

PN	DE4021917-A.
PD	16-JAN-1992.
PE	10-JUL-1990; 021917.
PF	10-JUL-1990; DE-021917.
PG	(BOEH) BOEHRINGER INGELHEI.
PH	Himmeler A, AdoIf G;
PI	WFI; 92-025485/04.
PJ	O-glycosylated alpha-interferon, used as medicament - isolated
PK	PT following secretion into conditioned medium of mammalian cells
PL	contg. a suitable expression plasmid.
PM	Disclosure; Fig 7(A-D); 24pp; German.
PN	PAD-CMW19 includes CMV promoter, modified DHER gene, splice donor
PO	acceptor sites, SV40 early adenylating signal, SV40 origin, and
PP	cloning site. The 5' non coding region of human IFNalpha 2c-e-
PQ	cDNA was replaced with the non coding region of human beta-globin
PR	using PCR. A PCR product of 0.64kb resulting from HindIII and
PS	restriction enzyme digests was isolated and ligated into appro-
PT	restriction enzyme sites of plasmid PAD-CMW19. E.coli HB101 w
PU	transformed with the ligation prods. and an expression plasmid
PV	PAD19B-IrN secreting O-glycosylated human IFN-alpha 2C was ide
PW	See also Q20764-66 and Q23517-29.
PX	Sequence 6062 BP:
QO	1577 A:
SO	1504 C: 1526 G: 1455 T

alignment_scores:		
Quality:	44.00	Length: 15
Ratio:	3.143	Gaps: 0
Percent similarity:	93.333	Percent Identity: 60.000

alignment block:

Align seq 1/1 to: Q20765 from: 1 to: 6062

3 LeuAlaIleArgLeuAsnGluArgTyrArgLeuAlaIleArgLeu 17  
|||||:::|||||:::|||||:::|||||  
1174 CTGGCCATTCACGTAAACAGAAGATTCCGCCTCAAGTTCCGGTTA 1218

seq name: N Geneseq 36:020732

```
seq_documentation_block:
ID Q20732 standard; DNA; 6062 BP.
AC Q20732;
DT 19-MAY-1992 (first entry)
DE PAD-CMV19 expression vector.
KW Interferon; IFN; O-glycosylation; ss.
```



PI Adolf G, Himmler A, Ahorn HJ, Kalsner I, Maurer-Fogy I;  
PT WPI; 92-056870/07.  
DR O-glycosylated alpha-interferon - used for treatment of  
PT viral of tumour diseases  
PS Disclosure: Fig 7 (A-D); 10app; English.  
CC A DNA sequence encoding IFN-alpha may be inserted into PAD-CMV19  
CC and the recombinant prod. used to transform cells of a  
CC multicellular organism, pref. a vertebrate. The cell supernatant  
CC is harvested and the O-glycosylated prod. recovered by known  
CC methods.  
CC See also Q20731-43 and Q20523-26.  
CC Sequence 6062 BP; 1580 A; 1506 C; 1528 G; 1448 T;  
SQ

alignment\_scores:  
Quality: 44.00 Length: 15  
Ratio: 3.143 Gaps: 0  
Percent Similarity: 93.33 Percent Identity: 60.000

alignment\_block:  
US-08-653-294-31 x Q20732 ..  
Align seg 1/1 to: Q20732 from: 1 to: 6062

3 LeuAlaIleArgLeuAsnGluArgTyrArgLeuAlaIleArgLeu 17  
|||||  
1174 CTGGCCATTCAGTAAACAGAGAGATTCGGCTCAAGTTCGGTTA 1218

seq\_name: N\_Genseq\_36:Q06283

seq\_documentation\_block:  
ID Q06283 standard; DNA; 6414 BP.  
AC Q06283;  
DT 29-JAN-1991 (first entry)  
DE PAD-CMV1 contg. a tumour Necrosis Factor-Binding Protein insert.  
KW Tumour necrosis factor binding protein; TNF-BP; TNF-receptor;  
KW PAD-CMV1; ss.  
OS Homo sapiens.  
PN EP-393438-A.  
PD 24-OCT-1990.  
PF 06-APR-1990; 106624.  
PR 21-APR-1989; DE-913101.  
PR 21-JUN-1989; DE-920282.  
PA (BOEH ) BOEHRINGER INGELHEIMINT.  
PI Hauptmann R, Himmler A, Maurer-Fogy I, Stratowa C;  
PT WPI; 90-321987/43.  
PT DNA encoding TNF binding protein and TNF- receptor - used in  
PT tumour treatment and to understand mechanisms to TNF action  
PS Disclosure; Fig 6(1-4); 51pp; German.  
CC PTNF-BP15 N(6282) is One of 30 positives clones in a screened cDNA  
CC library from induced TNF-induced fibrosarcoma cells. A TNF-BP had been  
CC isolated from the urine of patients with uraemia and probes/primers  
CC were constructed from the determined amino acid sequence.  
CC To produce a vector expressing a soluble form of TNF-binding  
CC protein, this plasmid was cut with XmnI, amplified by PCR and the  
CC amplified DNA cut with BamHI and EcoRI.  
CC The resulting 0.75 kb DNA fragment was inserted into pT7/T3 alpha-19  
CC (BRL) cut with the same enzymes to recover pTNF-BP. This was cut  
CC with BamHI and EcoRI, and the recovered fragment inserted into  
CC PAD-CMV1 to give the required plasmid pADTNF-BP.  
CC See also Q06282-Q06285.  
CC Sequence 6414 BP; 1679 A; 1565 C; 1554 G; 1616 T;  
SQ

alignment\_scores:  
Quality: 44.00 Length: 15  
Ratio: 3.143 Gaps: 0  
Percent Similarity: 93.33 Percent Identity: 60.000

alignment\_block:  
US-08-653-294-31 x Q06283 ..  
Align seg 1/1 to: Q06283 from: 1 to: 6414

3 LeuAlaileArgLeuAsnGluArgTyrArgLeuAlaileArgLeu 17  
 |||||:::|||||:::|||||:::|||||:::|||||  
 1531 CTGGCCATTACGTAAACAGAGATTCGGCTCAAGTCCGGTTA 1575

seq\_name: N\_Geneseq\_36:Q20766

seq\_documentation\_block:

ID Q20766 standard; DNA; 6414 BP.  
 AC Q20766; 21-APR-1992 (first entry)  
 DE PAD-CMV1.  
 KW Interferon; O-glycosylation; pCDM8; pSV2gptDHR20; ss.  
 FH Key Location/Qualifiers  
 FT promoter 1..21  
 FT /tag= a  
 FT /label= EBI-1733(Q22529)  
 FT /note= "start of CMV enhancer-promoter (from CDM8)"  
 FT enhancer 1..21  
 FT /tag= b  
 FT /label= EBI-1733(Q22529)  
 FT /note= "start of CMV enhancer-promoter (from CDM8)"  
 FT promoter 629..649  
 FT /tag= c  
 FT /label= T7\_promoter  
 FT misc\_feature 658..713  
 FT /tag= d  
 FT /label= multi-cloning\_site  
 FT /note= "HindIII-XbaI from EBI-1823, EBI-1829"  
 FT intron 714..1412  
 FT /tag= e  
 FT /note= "SV40 intron and poly-adenylation site (from CDM8)"  
 FT poly\_a\_signal 714..1412  
 FT /tag= f  
 FT /note= "SV40 intron and poly-adenylation site (from CDM8)"  
 FT 5'utr 1413..2310  
 FT /tag= g  
 FT /note= "5' non-coding region and promoter of hamster DHR gene (from pSV2gptDHR20)"  
 FT promoter 1413..2310  
 FT /tag= h  
 FT /note= "5' non-coding region and promoter of hamster DHR gene (from pSV2gptDHR20)"  
 FT exon 2311..2396  
 FT /tag= i  
 FT /label= DHR  
 FT /note= "hamster DHR: exon 1"  
 FT misc\_difference 2516  
 FT /tag= j  
 FT /label= mutation  
 FT /note= "A->T, disturbs PstI site"  
 FT exon 2701..3178  
 FT /tag= k  
 FT /label= DHR  
 FT /note= "exons 2-6"  
 FT misc\_difference 2707  
 FT /tag= l  
 FT /label= mutation  
 FT /note= "A->G, disturbing EcoRI site"  
 FT misc\_difference 3272..3273  
 FT /tag= m  
 FT /label= deletion  
 FT /note= "deletion between BglII and BamHI in DHR 3' non-coding region"  
 FT misc\_rna 3831  
 FT /tag= n  
 FT /note= "end of DHR gene (from pSV2gptDHR20)"  
 FT rep\_origin 3832..4169  
 FT /tag= o  
 FT /label= SV40\_ori  
 FT /note= "from SV2gptDHR20"

FT rep\_origin 4170..4648  
 FT /tag= p  
 FT /label= M13\_ori  
 FT /note= "from pBluescript SK+"  
 FT cds 4780..5640  
 FT /tag= q  
 FT /label= beta\_lactamase  
 FT misc\_feature 6395..6414  
 FT /tag= r  
 FT /label= EBI-1729(Q22528)  
 FT /note= "end of pBluescript vector sequence"  
 FT PN DE4021917-A.  
 FT PD 16-JAN-1992.  
 FT PE 10-JUL-1990; 021917.  
 FT PR 10-JUL-1990; DE-021917.  
 FT PA (BOEH ) BOEHRINGER INGELHEI.  
 FT PI Himmler A, Agolf G;  
 FT DR WPI; 92-025485/04.  
 FT PT O-glycosylated alpha-interferon, used as medicament - isolated following secretion into conditioned medium of mammalian cells  
 FT contg. a suitable expression plasmid  
 FT PS Disclosure; Fig 11(A-D); 24pp; German.  
 FT CC Plasmid pAD-CMV1 is formed from pCDM8, pSV2gptDHR20 and pBluescript SK+. The multi-cloning site is useful for the insertion of  
 FT CC heterologous DNA, e.g. IFN. The plasmid is replicable in E.coli.  
 FT CC See also Q20764-66 and Q22517-29.  
 FT SQ Sequence 6414 BP; 1677 A; 1563 C; 1557 G; 1617 T;

alignment\_scores:

Quality: 44.00 Length: 15  
 Ratio: 3.143 Gaps: 0  
 Percent Similarity: 93.333 Percent Identity: 60.000

alignment\_block:

US-08-653-294-31 x Q20766 ..

Align seg 1/1 to: Q20766 from: 1 to: 6414

3 LeuAlaileArgLeuAsnGluArgTyrArgLeuAlaileArgLeu 17  
 |||||:::|||||:::|||||:::|||||:::|||||  
 1531 CTGGCCATTACGTAAACAGAGATTCGGCTCAAGTCCGGTTA 1575

seq\_name: N\_Geneseq\_36:Q20733

seq\_documentation\_block:

ID Q20733 standard; DNA; 6414 BP.  
 AC Q20733;  
 DT 19-MAY-1992 (first entry)  
 DE PAD-CMV1 expression vector.  
 KW Interferon; IFN; O-glycosylation; ss.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..21  
 FT /tag= a  
 FT /note= "binding region of EBI-1733 (Q20525), start of CMV enhancer-promoter (from CDM8)"  
 FT enhancer 1..21  
 FT /tag= b  
 FT /note= "binding region of EBI-1733 (Q20525), start of CMV enhancer-promoter (from CDM8)"  
 FT promoter 1..21  
 FT /tag= c  
 FT /note= "binding region of EBI-1733 (Q20525), start of CMV enhancer-promoter (from CDM8)"  
 FT promoter 632..649  
 FT /tag= d  
 FT /label= T7\_promoter  
 FT misc\_feature 658..713  
 FT /tag= e  
 FT /label= multi-cloning\_site  
 FT /note= "HindIII-XbaI from EBI-1823, EBI-1829"  
 FT intron 714..1412  
 FT /tag= f

```

FT FT poly_a_signal /note= "SV40 intron and poly(A) site (from CDM8)"
FT FT 714..1412 /*tag= g
FT FT /note= "SV40 intron and poly(A) site (from CDM8)"
FT FT 1413..2310 /*tag= h
FT FT /note= "promoter and 5' non-coding region of
FT FT hamster DHFR gene (from pSV2gptDHFR20)"
FT FT 1413..2310 /*tag= i
FT FT /note= "promoter and 5' non-coding region of
FT FT hamster DHFR gene (from pSV2gptDHFR20)"
FT FT 2311..2396 /*tag= j
FT FT /label= DHFR
FT FT /note= "exon 1"
FT FT 2516 /*tag= k
FT FT /label= mutation
FT FT /note= "A -> T, disturbing the PstI site of
FT FT DHFR intron 1"
FT FT 2701..3178 /*tag= l
FT FT /label= DHFR
FT FT /note= "exons 2-6"
FT FT 2707 /*tag= m
FT FT /label= mutation
FT FT /note= "A -> G, disturbing the EcoRI site"
FT FT 3272..3273 /*tag= n
FT FT /note= "deletion between BglII and BamHI in
FT FT DHFR 3' non-coding region"
FT FT 3831 /*tag= o
FT FT /note= "end of DHFR gene (from pSV2gptDHFR20)"
FT FT 3832..4169 /*tag= p
FT FT /label= SV40_ori
FT FT /note= "from pSV2gptDHFR20"
FT FT 4170..4648 /*tag= q
FT FT /label= M13_ori
FT FT /note= "from pBluescript SK+"
FT FT 4780..5640 /*tag= r
FT FT /label= beta-lactamase
FT FT 6395..6414 /*tag= s
FT FT /note= "binding region of EBI-1729 (Q20524),
FT FT end of pBluescript vector"
FT FT
FT FT W09201055-A.
FT FT 23-JAN-1992.
FT FT 06-JUL-1991; E01266.
FT FT 10-JUL-1990; DE-021917.
FT FT 12-NOV-1990; DE-035877.
FT FT (BOEH ) BOEHRINGER INGELHEIM.
FT FT Adolf G, Himmeler A, Ahorn HJ, Kalsner I, Maurer-Poggy I;
FT FT WFI; 92-056870/07.
FT FT O-glycosylated alpha-interferon - used for treatment of
FT FT viral of tumour diseases
FT FT Disclosure: Fig 11 (A-E); 104pp; English.
FT FT This expression plasmid contains an intron sequence between the
FT FT multi-cloning sequence and the poly(A) site. Variations may be
FT FT made from pAD-CMV1, according to the amt. and position of introns
FT FT w.r.t. the multi-cloning site. For example, pAD-CMV19 (Q20732)
FT FT comprises only one intron between the CMV promoter and the multi-
FT FT cloning site. A DNA sequence encoding IFN-alpha may be then be
FT FT inserted into pAD-CMV19 and the recombinant prod. used to transform
FT FT cells of a multicellular organism, pref. a vertebrate. The cell
FT FT supernatant is harvested and the O-glycosylated prod. recovered by
FT FT known methods.
FT FT See also Q20731-43 and Q20523-26.

```

```

SQ Sequence 6414 BP; 1678 A; 1562 C; 1558 G; 1616 T;

alignment_scores:
  Quality: 44.00 Length: 15
  Ratio: 3.143 Gaps: 0
  Percent Similarity: 93.333 Percent Identity: 60.000

alignment_block:
  US-08-653-294-31 x Q20733 ..
  Align seg 1/1 to: Q20733 from: 1 to: 6414
  3 LeuAlaIleArgLeuAsnGluArgTyrArgLeuAlaIleArgLeu 17
  |||||||:|||||:|||||:|||||:|||||:|||||:|||||
  1531 CTGGCCATTCAGTAAACAGAGATTCGCCTCAGATCCGGTTA 1575

seq_name: N_Geneseq_36:Q43360

seq_documentation_block:
  ID Q43360 standard; DNA; 6623 BP.
  AC Q43360;
  DT 13-SEP-1993 (first entry)
  DE Sequence of circular plasmid pAHygcMV1
  KW PCR; primer: oligonucleotide.
  OS Synthetic.
  FH Key Location/Qualifiers
  FT Promoter 1..767
  FT /*tag= a
  FT /label= CMV
  FT 768..785
  FT /*tag= b
  FT /label= T7
  FT 794..854
  FT /*tag= c
  FT /label= polycloning site
  FT 854..1552
  FT /*tag= d
  FT /label= SV40 t intro
  FT /note= "and polyA signal"
  FT 1553..1736
  FT /*tag= e
  FT /label= Hamster DHFR gene
  FT 1737..2261
  FT /*tag= f
  FT /label= EBV ori p
  FT 2262..2856
  FT /*tag= g
  FT /label= HSV Thymidine kinase
  FT /note= "with polyA signal"
  FT 2857..3912
  FT /*tag= h
  FT /label= Hygromycin B phosphotransferase gene
  FT 3913..4161
  FT /*tag= i
  FT /label= HSV Thymidine Kinase
  FT 4162..6531
  FT /*tag= j
  FT /label= pBR322
  FT 6532..6623
  FT /*tag= k
  FT /label= linker sequence
  FT
  FT W09311257-A.
  FT 10-JUN-1993.
  FT 25-NOV-1992; E02718.
  FT 25-NOV-1991; DE-138621.
  FT (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
  FT Czeruilofsky AP, Himmeler A, Stratowa C, Weyer U;
  FT Lanche H, Schaefer R;
  FT WFI; 93-197073/24.
  FT Screening substances that modulate receptor-dependent signal
  FT transmission path - using test cells transformed with reporter
  FT gene and regulatory sequence sensitive to

```

PT inositol-1,4,5-tri:peptide and di:acetyl:glycerine(s)  
 PS Example: Pages 102-107; 170 pp; German.  
 CC Plasmids for the expression of genes or cDNA under the  
 CC transcription control of CMV promoters/enhancer and selectable  
 CC for hygromycin B resistance were constructed from expression  
 CC plasmids pAD-CMV1 and pAD-CMV2 (EP-A 393 438) and pHEBO (Sugden  
 CC et al 1985). Plasmid pAHyGCW1 contains the many features,  
 CC given in FT. Plasmid pAHyGCW2 differs from pAHyGCW1 only in  
 CC respect of the polycloning site. 'N' in q43360 denotes where the  
 CC sequence printed in the patent application is illegible.  
 SQ Sequence 6623 BP; 1651 A; 1701 C; 1626 G; 1640 T;

alignment\_scores:  
 Quality: 44.00 Length: 15  
 Ratio: 3.143 Gaps: 0  
 Percent Similarity: 93.333 Percent Identity: 60.000

alignment\_block:

US-08-653-294-31 x Q43360

Align seg 1/1 to: Q43360 from: 1 to: 6623

3 LeuAlaIleArgLeuAsnGluArgTyArgLeuAlaIleArgLeu 17  
 |||||.....:|||||:|||||:|||||:|||||  
 1671 CTGCCATTACGTAACAGAGAAGATTCGCTCAAGTTCGGTTA 1715

seq\_name: N\_Geneseq\_36:Q43361

seq\_documentation\_block:

ID Q43361 standard; DNA; 6630 BP.

AC Q43361;

DT 13-SEP-1993 (first entry)

DE Sequence of circular plasmid pAHyGCW2

KW PCR; primer: oligonucleotide.

OS Synthetic.

PN WO9311257-A.

PD 10-JUN-1993.

PF 25-NOV-1992; E02718.

PR 25-NOV-1991; DE138621.

PA (BOEH ) BOEHRINGER INGELHEIM INT GMBH.

PI Czernilofsky AP, Himmler A, Stratowa C, Weyer U;

PI Lamche H, Schaefer R;

DR WPI: 93-197073/24.

PT Screening substances that modulate receptor-dependent signal

PT transmission path - using test cells transformed with reporter

PT gene and regulatory sequence sensitive to

PT inositol-1,4,5-tri:peptide and di:acetyl:glycerine(s)

PS Example: Pages 107-112; 170 pp; German.

CC Plasmids for the expression of genes or cDNA under the

CC transcription control of CMV promoters/enhancer and selectable

CC for hygromycin B resistance were constructed from expression

CC plasmids pAD-CMV1 and pAD-CMV2 (EP-A 393 438) and pHEBO (Sugden

CC et al 1985). Plasmid pAHyGCW1 contains the many features,

CC given in FT. Plasmid pAHyGCW2 differs from pAHyGCW1 only in

CC respect of the polycloning site. 'N' in q43360 denotes where the

CC sequence printed in the patent application was illegible.

SQ Sequence 6630 BP; 1655 A; 1704 C; 1628 G; 1643 T;

alignment\_scores:

Quality: 44.00 Length: 15  
 Ratio: 3.143 Gaps: 0  
 Percent Similarity: 93.333 Percent Identity: 60.000

alignment\_block:

US-08-653-294-31 x Q43361

Align seg 1/1 to: Q43361 from: 1 to: 6630

3 LeuAlaIleArgLeuAsnGluArgTyArgLeuAlaIleArgLeu 17

|||||.....:|||||:|||||:|||||:|||||

1678 CTGCCATTACGTAACAGAGAAGATTCGCTCAAGTTCGGTTA 1722

seq\_name: N\_Geneseq\_36:X20551

seq\_documentation\_block:

ID X20551 standard; DNA; 6993 BP.

AC X20551;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of Treponema pallidum.

KW Treponema pallidum infection; syphilis; Borrelia infection; animal;

OS enzyme production; ds.

PN Treponema pallidum.

PN WO9859034-A2.

PD 30-DEC-1998.

PF 23-JUN-1998; U13041.

PR 24-JUN-1997; US-050667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fraser CM;

DR WPI: 99-081273/07.

PT New isolated Treponema pallidum nucleic acids - used to develop

PT products for the detection, diagnosis, characterisation, prevention

PT and therapy of T. pallidum infections, particularly syphilis

PS Claim 1; Page 483-487; 1150pp; English.

CC X20500-21243 represent polynucleotide sequences from the genome of

CC Treponema pallidum. The sequences can be used for detection,

CC diagnosis, characterisation, prevention and therapy for T. pallidum

CC infections, particularly syphilis. They can also be used for detecting

CC diseases related to Borrelia infections in animals, and for the

CC production of biosynthetic products such as enzymes.

SQ Sequence 6993 BP; 1914 A; 2019 C; 1404 G; 1647 T;

alignment\_scores:

Quality: 44.00 Length: 14  
 Ratio: 3.385 Gaps: 0  
 Percent Similarity: 92.857 Percent Identity: 64.286

alignment\_block:

US-08-653-294-31 x X20551

Align seg 1/1 to: X20551 from: 1 to: 6993

5 IleArgLeuAsnGluArgTyArgLeuAlaIleArgLeuAsn 18

|||||.....:|||||:|||||:|||||:|||||

6506 CTGCGCTTAACTCTCGAGAGACTGTCAATTCGTTGGAC 6547

seq\_name: N\_Geneseq\_36:Q34545

seq\_documentation\_block:

ID Q34545 standard; DNA; 589 BP.

AC Q34545;

DT 07-JUN-1993 (first entry)

DE -4aa IL-6 SSSS mutain.

KW Interleukin-6; mutant protein; pBgal/EK/cfIL-6; ss.

OS Homo sapiens.

PH Key

FT mat\_peptide

FT 1..543

FT /tag= a

FT /product= IL-6\_mutein

FT /note= "first 4 N-terminal amino acids are deleted

FT and Cys residues at positions 45, 51, 74

FT and 84 of wild-type IL-6 are substd. by

FT Ser"

FT WO9301212-A.

PN 21-JAN-1993.

PF 02-JUL-1992; U05612

PR 02-JUL-1991; US-724698.

PA (IMCL-) IMCLONE SYSTEMS INC.

PA (UYNC-) UNIV NORTH CAROLINA.

PI Fowlkes DM, Skelly SM, Snouwaert JN, Tackney CT;

DR WPI: 93-045433/05.

DR P-PSDB: R31996.

PT New cysteine-depleted interleukin-6 mutein - for promoting cell

PT differentiation (e.g. of B-cells), as antiinflammatory agents and

PT for treating thrombocytopoenia  
 CC Example 1; Fig 2; 118pp; English.  
 CC Plasmid pBgal/EK/cfil-6 contains a DNA sequence which encodes a  
 CC fusion protein comprising beta-galactosidase, followed by an  
 CC enterokinase cleavage site, which is, in turn, followed by a  
 CC synthetic IL-6 peptide sequence. The IL-6 mutin sequence is that  
 CC of native IL-6 except that the four cysteine residues, which occur  
 CC at positions 45, 51, 74 and 84 of the mature full-length IL-6  
 CC molecule are replaced by serine residues. A truncated 0.58kb  
 CC Cysteine-free IL-6 sequence lacking the first 4 N-terminal amino  
 CC acids can be isolated by EcoRII-HindIII digestion. The fragment was  
 CC used in the preparation of various IL-6 muteins, either full-length  
 CC or N-terminally truncated. It was found that restoring the last two  
 CC Cys residues (at positions 74 and 84) resulted in a mutin which  
 CC retained IL-6 activity.  
 SQ Sequence 589 BP; 193 A; 145 C; 134 G; 117 T;

alignment\_scores:  
 Quality: 43.00 Length: 14  
 Ratio: 3.909 Gaps: 0  
 Percent Similarity: 78.571 Percent Identity: 57.143

alignment\_block:

US-08-653-294-31 x Q34545

Align seg 1/1 to: Q34545 from: 1 to: 589

3 LeuAlaIleArgLeuAnGluArgTyrArgLeuAlaIleArg 16  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 530 CTCGCTCAGTAAATGATAGTAGTACCGAGCTCGAATTCGT 571

seq\_name: N\_Geneseq\_36:Q89359

seq\_documentation\_block:

ID Q89359 standard; DNA; 589 BP.  
 AC Q89359;  
 DT 07-JUL-1995 (first entry)  
 DE Human interleukin-6 -(Ala1), Ser45, Ser51, Ser74, Ser84 DNA.  
 KW Homo sapiens.  
 OS Homo sapiens.  
 FH Key  
 FT mat\_peptide 1..543  
 FT Location/Qualifiers  
 FT /\*tag= a  
 PS US5359034-A.  
 PD 25-OCT-1994.  
 PF 02-JUL-1991: 724698.  
 PR 02-JUL-1991: US-724698.  
 PR 02-JUL-1992: US-907710.  
 PA (IMCL-) IMCLONE SYSTEMS INC.  
 PA (UYNC-) UNIV NORTH CAROLINA.  
 PI Fowlkes DM, Skelly SM, Snouwaert JN, Tackney CT;  
 DR WPI: 94-341061/42.  
 DR P-PSDB: R72211.  
 PT IL-6 Muteins in which the first two cysteine residues are  
 PT replaced by serine residues - have good activity comparable to  
 PT that of native IL-6  
 PS Example 1; Fig 2; 37pp; English.

CC Q89359 encodes R72211 the human interleukin-6 (IL-6) mutin  
 CC -(Ala1), Ser 45, Ser 51, Ser 74, Ser 84; where the native  
 CC IL-6 cysteine residues in positions 45, 51, 74 and 84 have  
 CC been replaced by serine residues, and the Ala residue in  
 CC position 1 is missing.  
 SQ Sequence 589 BP; 193 A; 145 C; 134 G; 117 T;

alignment\_scores:  
 Quality: 43.00 Length: 14  
 Ratio: 3.909 Gaps: 0  
 Percent Similarity: 78.571 Percent Identity: 57.143

alignment\_block:

US-08-653-294-31 x Q89359

Align seg 1/1 to: Q89359 from: 1 to: 589

3 LeuAlaIleArgLeuAnGluArgTyrArgLeuAlaIleArg 16  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 530 CTCGCTCAGTAAATGATAGTAGTACCGAGCTCGAATTCGT 571

seq\_name: N\_Geneseq\_36:T39892

seq\_documentation\_block:

ID T39892 standard; DNA; 589 BP.  
 AC T39892;  
 DT 10-DEC-1996 (first entry)  
 DE Coding sequence for -4aa SSCC mutin of human Interleukin-6.  
 KW Interleukin-6; IL-6; mutin; proliferation; differentiation; therapy;  
 KW T cell; megakaryocyte; multipotent haematopoietic progenitor cell; BSF-2;  
 KW anti-inflammatory agent; thrombocytopaenia; bone marrow transplantation;  
 KW interferon-beta-2; B-cell stimulation factor-2; 26 kDa protein; IFN-B2;  
 KW B-cell hybridoma/plasmacytoma growth factor; HPGF; HGF; HSF;  
 KW hepatocyte stimulating factor; ss.  
 OS Synthetic.  
 FH Key  
 FT Location/Qualifiers  
 FT 1..546  
 FT /\*tag= a  
 FT /product= -4 aa interleukin-6 SSCC  
 FT misc\_feature 1..72  
 FT /\*tag= b  
 FT /\*note= optionally deleted"

US545537-A.  
 PD 13-AUG-1996.  
 PF 02-JUL-1991: 724698.  
 PR 02-JUL-1991: US-724698.  
 PR 02-JUL-1992: US-907710.  
 PR 10-MAR-1994: US-209182.  
 PA (UYPR-) UNIV PRINCETON.  
 PI Fowlkes DM, Skelly SM, Snouwaert JN, Tackney CT;  
 DR WPI: 96-383669/38.  
 DR P-PSDB: W05586.  
 PT New nucleic acid encoding mutant IL-6 with two Cys residues replaced  
 PT - and opt. N-terminal deletion, can be produced in high yield and  
 PT are more active than native protein, useful e.g. as  
 PT immuno-therapeutic or anti-inflammatory agents  
 PS Example 2; Column 31-34; 41pp; English.

CC This sequence represents the coding sequence for a interleukin-6 (IL-6)  
 CC mutin of the invention. IL-6 is also known under numerous synonyms,  
 CC including interferon-beta-2 (IFN-B2), B-cell stimulation factor (BSF-2),  
 CC B-cell hybridoma/plasmacytoma growth factor (HPGF or HGF), 26 kDa  
 CC protein, and hepatocyte stimulating factor (HSF). The mutin IL-6  
 CC proteins can be used for regulating cells, to stimulate proliferation and  
 CC differentiation of B or T cells, megakaryocytes or multipotent  
 CC haematopoietic progenitor cells. The muteins can also induce acute phase  
 CC proteins in liver cell. The proteins can be used as immunotherapeutic  
 CC and anti-inflammatory agents, in the treatment of thrombocytopaenia, and  
 CC to treat patients undergoing chemotherapy or bone marrow transplantation.  
 CC By eliminating 2 of the Cys residues, purification and isolation of IL-6  
 CC is simplified as only one Cys-Cys disulphide bond can form during  
 CC oxidation. Due to this, yields of the mutin proteins are improved in  
 CC comparison to wild type IL-6. By removing the hydrophobic N-terminus,  
 CC the yield can be further improved. The IL-6 muteins of the invention  
 CC which have 22 N-terminal residues deleted have activity 250-400 times  
 CC that of the wild type IL-6.

SQ Sequence 589 BP; 193 A; 145 C; 134 G; 117 T;

alignment\_scores:  
 Quality: 43.00 Length: 14  
 Ratio: 3.909 Gaps: 0  
 Percent Similarity: 78.571 Percent Identity: 57.143

alignment\_block:

US-08-653-294-31 x T39892

Align seg 1/1 to: T39892 from: 1 to: 589

3 LeuAlaIleArgLeuAsnGluArgTyrArgLeuAlaIleArg 16  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 530 CTCTGCTCAGATGTAATGATAGGTACCGAGCTCGAATCGT 571

seq\_name: N\_Geneseq\_36:v37110

seq\_documentation\_block:

ID V37110 standard; DNA; 846 BP.

AC V37110;

DT 04-SEP-1998 (first entry)

DE Antibiotic resistance gene blaZ for beta-lactams.

KW Detection; bacterial antibiotic resistance gene; bacteria;

KW fungal species; identification; beta-lactam; ds.

OS Enterococcus sp.

PN W09820157-A2.

PD 14-MAY-1998.

PF 04-NOV-1997; CA0829.

PR 04-NOV-1996; US-743637.

PA (IDI-) IDI INFECTION DIAGNOSTIC INC.

PI Bergeron MG, Ouellette M, Picard FJ, Roy PH;

PT WPI; 98-286967/25.

PS Use of oligo:nucleotide primers and probes - for detection,

PT identification and quantification of bacteria, fungi and bacterial

PT antibiotic resistance gene(s)

PS Claim 19: Page 108; 167pp; English.

CC The present sequence represents an antibiotic resistance gene blaZ  
 CC (resistant to beta-lactams), and was used to design PCR primers  
 CC V37050-51. The specification describes the use of probes and/or  
 CC amplification primers which are specific, ubiquitous and sensitive for  
 CC determining the presence and amount of nucleic acids from a bacterial  
 CC antibiotic resistance gene and specific bacterial and fungal species in  
 CC any sample suspected of containing the bacterial or fungal nucleic acids,  
 CC where each of the nucleic acid or variant or part comprises a selected  
 CC target region hybridisable with the probes or primers. The method of  
 CC use comprises contacting the sample with the probes or primers and  
 CC detecting the presence of hybridised probes or amplified products as an  
 CC indication of the presence of the specific bacterial or fungal species  
 CC and bacterial antibiotic resistance genes. The methods and products can  
 CC be used to detect and identify the bacterial and fungal species and  
 CC genera and determine the bacterial resistance to antibiotics.

CC Sequence 846 BP; 356 A; 99 C; 135 G; 256 T;

alignment\_scores:

Quality: 43.00

Ratio: 3.583

Percent Similarity: 100.000

Percent Identity: 58.333

alignment\_block:

US-08-653-294-31 x V37110/rev ..

Align seg 1/1 to reverse of: V37110 from: 1 to: 846

8 AsnGluArgTyrArgLeuAlaIleArgLeuAsnGlu 19

||||:||||| |||:||||| |||:||||| |||:|||||

753 AACATAGGTTCAGATGTCCTTGGCCCTTAGGATAACAAA 718

seq\_name: N\_Geneseq\_36:T07329

seq\_documentation\_block:

ID T07329 standard; CDNA; 1391 BP.

AC T07329;

DT 05-JUL-1996 (first entry)

DE Human calcium signal-modulating cyclophilin ligand coding sequence.

KW Human calcium signal-modulating cyclophilin ligand; DNA probe;

KW calcium-activated pathway modulator drug screening; diagnosis;

KW treatment; contraception; infertility; learning disorder;

KW memory disorder; Escherichia coli; Saccharomyces cerevisiae; CHO;

KW COS; Jurkat; Chinese hamster ovary; H9c2(2-1); ds.

OS Homo sapiens.

FH Key

Location/Qualifiers

FT cds 37..927  
 ET /\*tag= a  
 PN W09535501-A1.  
 PD 28-DEC-1995.  
 PF 16-JUN-1995; U07752.  
 PR 17-JUN-1994; US-261662.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Bram RJ, Crabtree GR;  
 DR WPI; 96-058511/06.  
 DR P-PSDB; R87038.  
 PT Calcium signal-modulating cyclophilin, and DNA encoding it - useful  
 PT for diagnostic purposes, and for identifying agents for modulating  
 PT the calcium-induced activation pathway.  
 PS Claim 1: Page 20-21; 33pp; English.  
 CC This sequence encodes human calcium signal-modulating cyclophilin  
 CC ligand (CAML). It may be used as a probe to identify CAML genes in  
 CC hosts other than human, in diagnosis, to screen agents for their  
 CC effect on CAML expression, to provide antisense sequences to  
 CC investigate cellular responses to external agents, to produce  
 CC recombinant CAML protein, and to express a fragment of the CAML  
 CC protein to act as a dominant negative. Cells which overexpress  
 CC CAML may also be used to screen putative modulators of the  
 CC calcium-activated pathway for their effects on expression of one or  
 CC more genes under regulatory control of the pathway. Host cells for  
 CC expression of recombinant CAML include Escherichia coli,  
 CC Saccharomyces cerevisiae, CHO and COS cells, while host cells used  
 CC to investigate the role of CAML are usually mammalian cells, and  
 CC particularly human cells such as Jurkat T-lymphocytes and H9c2(2-1).  
 CC Targets for treatment mediated by CAML activity can include  
 CC therapeutic contraception, infertility, learning and memory  
 CC disorders and the like.  
 CC Sequence 1391 BP; 417 A; 280 C; 332 G; 362 T;

alignment\_scores:

Quality: 43.00

Ratio: 3.071

Percent Similarity: 93.333

Percent Identity: 46.667

alignment\_block:

US-08-653-294-31 x T07329/rev ..

Align seg 1/1 to reverse of: T07329 from: 1 to: 1391

6 ArgLeuAsnGluArgTyrArgLeuAlaIleArgLeuAsnGluArg 20

||||:||||| |||:||||| |||:||||| |||:|||||

1194 AAGTTAGAGGAAAGGTACAGATAGGAATTCGTATTAACAGAAA 1150

OM of: US-08-653-294-31 to: EST:\* out\_format : pfs  
 Date: Feb 8, 2000 6:23 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
 -MODEL=framet.p2n.model -DEV=xlp  
 -Q=Cgml\_1/USPTO\_spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.2  
 -DB=EST -QFWT=fastac -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500  
 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOR=6.000  
 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOR=6.000  
 -DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi  
 -LIST=45 -DOALIGN=200 -THR.SCORE=pct -ALIGN=15 -MODE=LOCAL  
 -OUTFWT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

#### Search information block:

Query: US-08-653-294-31  
 Query length: 20  
 Database: EST:\*  
 Database sequences: 4538634  
 Database length: 1887831982  
 Search time (sec): 7600.090000

#### score\_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
gb_est27:AW015295	- 52.00	158.16	4.81	527	AW015295 UI-H-BIO-aap-d-02-0-UI-s
gb_est28:AW149243	- 52.00	156.46	5.99	473	AW149243 ti27g11.xl NCI_CGAP_Ki
gb_est32:AW1762057	- 52.00	156.45	6.00	534	AW1762057 w53a10.xl NCI_CGAP_Co
gb_est27:AW1437378	- 51.00	153.27	9.12	525	AW1437378 fb30h04.xl Zebrafish W
gb_est27:AW1416006	- 51.00	153.18	9.12	531	AW1416006 fb30h04.xl Zebrafish W
gb_est40:AW147227	- 48.00	150.16	13.43	246	AW147227 dal6c03.v1 normalized
gb_est40:AW147713	- 48.00	150.16	13.43	246	AW147713 dal6c03.v1 normalized
gb_est33:AV074366	- 48.00	148.87	15.86	287	AV074366 AV074366 Mus musculus
gb_est22:AW1070842	- 48.00	147.32	19.33	345	AW1070842 UI-R-C2-mv-g-10-0-UI-s
gb_est37:AW1998122	- 48.00	142.98	33.74	579	AW1998122 701672303 A1. thaliana,
gb_est37:AW1997815	- 47.00	141.33	41.71	484	AW1997815 701554353 A1. thaliana,
gb_est26:AW1328939	- 47.00	140.94	43.84	507	AW1328939 a8a03ne.fl Neurospora
gb_gss14:AW0514060	- 47.00	140.94	43.84	507	AW0514060 HS_5184_A1_E05.SP6E RF
gb_gss13:AW0500194	- 47.00	140.52	46.27	533	AW0500194 V53C7 mTo-3xHA/lacZ In
gb_est41:AV0266473	- 46.00	143.03	33.51	271	AV0266473 AV266473 RIKEN full-1e
gb_est37:AW1998122	- 46.00	141.38	41.41	330	AW1998122 701672303 A1. thaliana,
gb_gss13:AW0593350	- 46.00	136.64	76.11	581	AW0593350 HS_5462_A2_D03.SP6E RF
gb_est32:AW1757858	- 45.00	138.94	56.64	303	AW1757858 ea19b02.y1 Elmeria M5-
gb_est32:AW1757463	- 45.00	137.00	72.67	382	AW1757463 ea05d10.y1 Elmeria M5-
gb_est32:AW1755661	- 45.00	134.96	94.37	487	AW1755661 ea07b04.y1 Elmeria M5-
gb_gss11:AW0305795	- 45.00	132.90	123.00	623	AW0305795 HS_2050_A1_B12.MR CIT
gb_gss14:AW051248	- 45.00	131.34	150.17	750	AW051248 HS_5574_B2_E09.SP6 RPO
gb_gss4:AW0691068	- 45.00	130.33	170.95	846	AW0691068 nbxb0086p18r CUGI Rice
gb_gss15:AW0591361	- 44.50	131.86	140.48	584	AW0591361 HS_5410_B2_E01.T7A RPO
gb_est34:AW139866	- 44.00	136.15	81.00	290	AW139866 AV139866 Mus musculus
gb_est1:DW1547	- 44.00	134.99	93.99	333	AW1547 RCC0652A Rice callus Or
gb_est33:AW209827	- 44.00	134.79	96.42	341	AW209827 AV209827 RIKEN full-1e
gb_est19:AW173968	- 44.00	133.67	111.41	390	AW173968 ac35a12.s1 Stragene
gb_gss10:AW0215334	- 44.00	132.91	122.82	427	AW0215334 HS_2259_B2_H06.MR CIT
gb_est14:AW0454364	- 44.00	132.83	124.06	431	AW0454364 MBFACX0F1173 Bruglia m
gb_est24:AW1103506	- 44.00	131.35	149.94	514	AW1103506 EST12795 Normalized
gb_gss11:AW0334845	- 44.00	130.21	173.60	589	AW0334845 HS_5018_A1_F09.T7 RPT
gb_est32:AW1738375	- 44.00	130.18	174.23	591	AW1738375 608049A10.xl 606 - Ead
gb_est19:AW263821	- 44.00	130.00	178.36	604	AW263821 LD07093.3prime LD Dros
gb_est28:AW1512344	- 44.00	130.00	178.36	607	AW1512344 LD44186.Sprime LD Dros
gb_est28:AW1543580	- 44.00	129.96	179.31	607	AW1543580 SD10377.5prime SD Dros
gb_est5:N32286	- 44.00	129.86	181.54	614	N32286 YX29d12.s1 Soares meland
gb_est30:AW164944	- 44.00	129.36	193.65	652	AW164944 486099E09.xl 486 - lea
gb_gss1:AW008484	- 44.00	128.43	218.04	728	AW008484 Homo sapiens genomic D
gb_gss1:AW008081	- 44.00	127.11	258.24	852	AW008081 Drosophila melanogaste
gb_gss6:AW0892181	- 44.00	124.27	371.61	1195	AW0892181 HS_5339_B1_F08.SP6 RF
gb_est30:C95817	- 43.50	126.78	269.34	734	C95817 C95817 Marchantia polym

gb\_est25:AW1287354 + 43.00 134.58 99.07 240 ! AW1287354 qv02b02.xl NCI\_CGAP  
 gb\_est37:AW013132 + 43.00 134.24 103.52 250 ! AW013132 SRN03SKS Winter f10  
 gb\_est25:AW1279828 - 43.00 132.77 125.05 298 ! AW1279828 qm21e02.xl NCI\_CGAP  
 seq\_name: gb\_est37:AW015295  
 seq\_documentation\_block:  
 LOCUS AW015295 427 bp mRNA EST 10-SEP-1999  
 DEFINITION UI-H-BIO-aap-d-02-0-UI-s1 NCI\_CGAP\_Sub1 Homo sapiens cDNA clone  
 IMAGE:2709939 3', mRNA sequence.  
 ACCESSION AW015295  
 VERSION AW015295.1 GI:5864052  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 427)  
 NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 COMMENT  
 On May 18, 1998 this sequence version replaced gi:3138601.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Oligo-dT track not found. Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: NCI\_CGAP clone distribution  
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: M13 Forward  
 POLYA-No.

#### FEATURES

Location/Qualifiers  
 1..427  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2709939"  
 /lab\_host="NCI\_CGAP\_Sub1"  
 /note="Vector: pYT3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The  
 NCI\_CGAP\_Sub1 library is a subtracted library derived from  
 BI. BI constitutes a mixture of 21 normalized or  
 subtracted NCI\_CGAP libraries: NCI\_CGAP\_Co4,  
 NCI\_CGAP\_P22, NCI\_CGAP\_P18, NCI\_CGAP\_Co10,  
 NCI\_CGAP\_Co16, NCI\_CGAP\_Kid5, NCI\_CGAP\_Kid12,  
 NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid11, NCI\_CGAP\_Lym2,  
 NCI\_CGAP\_Br2, NCI\_CGAP\_Co8, NCI\_CGAP\_CLL1, NCI\_CGAP\_Le12,  
 NCI\_CGAP\_Br23, NCI\_CGAP\_Lu5, NCI\_CGAP\_Lu24,  
 NCI\_CGAP\_Lu19, NCI\_CGAP\_G4, NCI\_CGAP\_G6, NCI\_CGAP\_Br25.  
 These 21 libraries were pooled and a single-stranded DNA  
 preparation of the resulting mixture was used as a tracer  
 in a subtractive hybridization with a driver whose  
 composition is detailed below: NCI\_CGAP\_Kid3 pool 1 LLAM  
 3334-3337, 3682-3683, 3798-3803 (IMAGE Cloneds)  
 1322376-1323911, 1456008-1456775, 1500552-1502855)  
 NCI\_CGAP\_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778  
 (IMAGE Cloneds 1323912-1325831, 1471368-1472903,  
 1492104-1493255) NCI\_CGAP\_Lu5 pool 1 LLAM 3575-3582,  
 3851-3854 (IMAGE Cloneds 1414920-1417991,  
 1520904-152439) NCI\_CGAP\_G4 pool 1 LLAM 3164-3167,  
 3716-3720, 3733-3735 (IMAGE Cloneds 1257096-1258631,  
 1458064-1470983, 1475592-1476743) NCI\_CGAP\_P22 pool 1  
 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Cloneds  
 985608-986759, 1101192-1101959, 1217928-1220615)  
 NCI\_CGAP\_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE  
 Cloneds 1057416-1061255, 1144584-1145351). The resulting  
 subtracted library contained 530,000 recombinants.  
 Subtraction was performed as previously described.  
 [Bonaldo, Lennon & Soares (1996): Normalization and  
 Subtraction: Two Approaches TO Facilitate Gene Discovery.  
 Genome Research 6, 791-806.]

TAG\_LIB=NCI\_CGAP\_Pr22  
TAG\_TISSUE=prostate  
TAG\_SEQ=AAAGTG

BASE COUNT 61 a 148 c 161 g 57 t  
ORIGIN

alignment\_scores:  
Quality: 52.00 Length: 18  
Ratio: 3.714 Gaps: 0  
Percent Similarity: 77.778 Percent Identity: 66.667

alignment\_block:  
US-08-653-294-31 x AW015295/rev ..

Align seg 1/1 to reverse of: AW015295 from: 1 to: 427

2 ArgLeuAlaIleArgLeuAsnGluArgTyArgLeuAlaIleArgLeuAs 18  
||||:|||||  
89 CCGGTGGCCATCCGGCGAATGAGCGCTACCGGCTGCTGGCGCGCAGCAC 40

18 nGlu 19  
:|  
39 CGAG 36

seq\_name: gb\_est38:AI492423

seq\_documentation\_block: 523 bp mRNA EST 30-MAR-1999  
LOCUS AI492423  
DEFINITION t127g11.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2131748 3',  
mRNA sequence.

ACCESSION AI492423  
VERSION AI492423.1 GI:4393426  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 523)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On May 7, 1998 this sequence version replaced gi:3121386.  
Contact: Robert Strausberg, Ph.D.

Email: Robert.Strausberg@nih.gov  
Tel: (301) 496-1550  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 731 Std Error: 0.00

Seq primer: -40UP from Gibco  
High quality sequence stop: 453.

FEATURES  
Location/Qualifiers

1..523  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2131748"  
/clone\_lib="NCI\_CGAP\_Kid11"

/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library

BASE COUNT 97 a 162 c 152 g 112 t  
ORIGIN

alignment\_scores:  
Quality: 52.00 Length: 18  
Ratio: 3.714 Gaps: 0  
Percent Similarity: 77.778 Percent Identity: 66.667

alignment\_block:  
US-08-653-294-31 x AI492423/rev ..

Align seg 1/1 to reverse of: AI492423 from: 1 to: 523

2 ArgLeuAlaIleArgLeuAsnGluArgTyArgLeuAlaIleArgLeuAs 18  
||||:|||||  
72 CCGGTGGCCATCCGGCGAATGAGCGCTACCGGCTGCTGGCGCGCAGCAC 23

18 nGlu 19  
:|  
22 CGAG 19

seq\_name: gb\_est32:AI762057

seq\_documentation\_block: 524 bp mRNA EST 24-JUN-1999  
LOCUS AI762057  
DEFINITION w153a10.x1 NCI\_CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2393946 3',  
similar to contains TARI.tl MER22 MER22 repetitive element ;, mRNA  
sequence.

ACCESSION AI762057  
VERSION AI762057.1 GI:5177724  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 524)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187034.  
Contact: Robert Strausberg, Ph.D.

Email: Robert.Strausberg@nih.gov  
Tel: (301) 496-1550  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 469.

FEATURES  
Location/Qualifiers

1..524  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2393946"  
/clone\_lib="NCI\_CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"

/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Col10 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs



```

from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo.
BASE COUNT      63 a   193 c   190 g   78 t
ORIGIN

alignment_scores:
  Quality:      52.00      Length:      18
  Ratio:        3.714      Gaps:        0
Percent Similarity: 77.778      Percent Identity: 66.667

alignment_block:
US-08-653-294-31 x AI762057/rev ..
Align seg 1/1 to reverse of: AI762057 from: 1 to: 524

2 ArgLeuAlaileArgLeuAsnGluArgTyrArgLeuAlaileArgLeuAs 18
|||||
64 CGCGTGGCCATCCGGCGAATGAGCGCTACCGCTGCTGCGCGCAGCAC 15
18 nGlu 19
14 CGAG 11

seq_name: gb_est27:AI437378

seq_documentation_block:
LOCUS      AI437378      525 bp      mRNA      09-MAR-1999
DEFINITION      fb30h04.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA 3' similar to
                SW:PTEN_HUMAN O00633 PUTATIVE PROTEIN-TYROSINE PHOSPHATASE PTEN ;
                mRNA sequence.
ACCESSION      AI437378
VERSION        AI437378.1 GI:4285423
KEYWORDS
SOURCE
ORGANISM
Danio rerio
Eukaryota; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasbora; Danio.
REFERENCE
AUTHORS
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
On May 18, 1998 this sequence version replaced gi:3138276.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Possible reversed clone: similarity on wrong strand
Seq primer: T7 ET from Amersham
High quality sequence stop: 426.
Location/Qualifiers
1..525
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"

FEATURES
source

```

```

/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'PGACTAGTCTAGATCGGCGCCGCGCTTTTITTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
BASE COUNT      95 a   129 c   119 g   179 t   3 others
ORIGIN

alignment_scores:
  Quality:      51.00      Length:      17
  Ratio:        3.188      Gaps:        0
Percent Similarity: 94.118      Percent Identity: 52.941

alignment_block:
US-08-653-294-31 x AI437378 ..
Align seg 1/1 to: AI437378 from: 1 to: 525

1 TyrArgLeuAlaileArgLeuAsnGluArgTyrArgLeuAlaileArgLe 17
|||||
245 TACAATTTGACCTTAAGTTGGTGGAATAATATCGTTAGCCTTATCTTT 294
17 u 17
295 G 295

seq_name: gb_est27:AI416006

seq_documentation_block:
LOCUS      AI416006      531 bp      mRNA      09-FEB-1999
DEFINITION      fb30h04.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
                SW:PTEN_HUMAN O00633 PUTATIVE PROTEIN-TYROSINE PHOSPHATASE PTEN ;
                mRNA sequence.
ACCESSION      AI416006
VERSION        AI416006.1 GI:4259510
KEYWORDS
SOURCE
ORGANISM
Danio rerio
Eukaryota; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasbora; Danio.
REFERENCE
AUTHORS
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
On May 7, 1998 this sequence version replaced gi:3119070.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

```

Email: zbrafish@watson.wustl.edu  
 CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
 Missouri (web address: www.genomesystems.com) (email contact:  
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
 (web address: www.resgen.com) (email contact: info@resgen.com) and  
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:  
 www.rzpd.de)

Seq primer: T3 ET from Amerham  
 High quality sequence stop: 485.

## FEATURES

source

1. .531  
 /db\_xref="taxon:7955"  
 /clone\_lib="zebrafish WashU MPIMG EST"  
 /sex="mixed"  
 /tissue\_type="26 somite embryos, adult livers, shield  
 stage embryos"  
 /lab\_host="X11-blue MRF"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
 strand cDNA was primed with a Not I - oligo(dT)15 primer  
 [5'pGACTAGTCTAGATCGAGCGCGCCCTTTTCTTTTCTTTT3'];  
 double-stranded cDNA was ligated to Sal I adaptors (BRL),  
 digested with Not I and cloned into the Not I and Sal I  
 sites of the pSPORT1 vector (BRL). Library was constructed  
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
 analysis were selected following oligonucleotide  
 hybridization fingerprinting of arrayed clones from  
 zebrafish late somitogenesis (26 ss), adult liver or  
 embryonic shield stage (5.6 h) libraries. Fingerprint  
 data were used to computationally cluster cDNAs, and a  
 single cDNA from each cluster was chosen for sequencing.  
 In some cases multiple members of the same cluster were  
 sequenced to assess clustering parameters or single clones  
 were sequenced additional times to assess quality  
 control."

BASE COUNT 158 a 119 c 144 g 110 t  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 17  
 Ratio: 3.188 Gaps: 0  
 Percent Similarity: 94.118 Percent Identity: 52.941

alignment\_block:  
 US-08-653-294-31 x AI416006/rev ..  
 Align seg 1/1 to reverse of: AT416006 from: 1 to: 531  
 1 TyrArgLeuAlaIleArgLeuAsnGluArgTyrArgLeuAlaIleArgLe 17  
 |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||  
 407 TACAATTGACCTTAAGTTGGGTGAAATATCGGTAGCTTACTTT 358

17 u 17  
 |  
 357 G 357

seq\_name: gb\_est40:AW147227

seq\_documentation\_block:  
 LOCUS AW147227 235 bp mRNA EST 30-NOV-1999  
 DEFINITION da16c03.x1 normalized Xenopus laevis gastrula Xenopus laevis cDNA  
 clone XENOPUS\_SOURCE\_ID:xlinga001n05 3', mRNA sequence.  
 ACCESSION AW147227  
 VERSION AW147227.1 GI:6195123  
 KEYWORDS EST,  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
 Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;

## REFERENCE

## AUTHORS

Xenopus  
 1 (bases 1 to 235)  
 Johnson,S.L., Blumberg,B., Song,J., Clifton,S., Hillier,L.,  
 Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y.,  
 Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,  
 Waterston,R. and Wilson,R.  
 WashU Xenopus EST project, 1999  
 Unpublished (1999)

## TITLE

## JOURNAL

## COMMENT

On Dec 20, 1995 this sequence version replaced gi:1135577.  
 Other\_ESTs: da16c03.y1  
 Contact: Stephen L. Johnson/WashU Xenopus EST project, 1999  
 WashU Xenopus EST project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Library constructed by Bruce Blumberg  
 Library normalized by Jihwan Song  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clone distribution information for  
 this library can be found through Research Genetics, visit their  
 web page at: http://www.resgen.com/  
 Seq primer: -40UP from Gibco.

## FEATURES

source

1. 235  
 Location/Qualifiers  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="XENOPUS\_SOURCE\_ID:xlinga001n05"  
 /clone\_lib="normalized Xenopus laevis gastrula"  
 /tissue\_type="gastrula (stages 10.5, 11.5 mixed)"  
 /lab\_host="Top 10 F"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal  
 parts from stage 10.5 and stage 11.5 gastrulae).  
 EcoRI-XhoI cut cDNA was then ligated into Unizap-XR  
 (Stratagene) with EcoRI at the 5' end and XhoI at the 3'  
 end. SS-library phagemids were prepared by mass excision  
 from the original library and normalized by hybridization  
 to biotinylated driver (prepared from the same library by  
 PCR) to cot-omega of 11. After removal of hybrids and  
 excess driver by streptavidin sepharose chromatography,  
 the ss-phagemids were made double stranded and  
 electroporated into Top-10 F'. Original library  
 constructed by Bruce Blumberg (Cho et al. 1991 Cell 67,  
 1111-1120). Normalized by Jihwan Song (Song, Cho and  
 Blumberg, unpublished)."  
 70 a 52 c 45 g 68 t

alignment\_scores:  
 Quality: 48.00 Length: 12  
 Ratio: 4.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:  
 US-08-653-294-31 x AW147227/rev ..

Align seg 1/1 to reverse of: AW147227 from: 1 to: 235

8 AsnGluArgTyrArgLeuAlaIleArgLeuAsnGlu 19  
 ::|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||  
 214 TCTGAGCGCTACAGATGGCCATACGTGTGTGACACAG 179

seq\_name: gb\_est40:AW147713

seq\_documentation\_block:  
 LOCUS AW147713 246 bp mRNA EST 30-NOV-1999  
 DEFINITION da16c03.y1 normalized Xenopus laevis gastrula Xenopus laevis cDNA  
 clone XENOPUS\_SOURCE\_ID:xlinga001n05 5', mRNA sequence.  
 ACCESSION AW147713  
 VERSION AW147713.1 GI:6195609

LOCUS	AV074366	287 bp	EST	24-JUN-1999
DEFINITION	AV074366 Mus musculus stomach C57BL/6J adult Mus musculus cDNA clone 2210008P05, mRNA sequence.			
ACCESSION	AV074366			
VERSION	AV074366.1	GI:5194194		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 287)			
AUTHORS	Carlinici,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akhira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomimaga,N., Watanabe,S., Yagane,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.			
TITLE	RIKEN Mouse ESTs			
JOURNAL	Unpublished (1999)			
COMMENT	On Dec 20, 1995 this sequence version replaced gi:1134354.			
	Contact: Chie Owa			
	Genome Science Laboratory			
	RIKEN			
	3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan			
	Tel: 81-298-36-9145			
	Fax: 81-298-36-9098			
	Email: genome.res@rcl.riken.go.jp			

Enthalpy	genome.res.ctc.riken.go.jp
Thermodynamic stability	Thermodynamic stability and thermocatalysis of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing	A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site	(http://genome.rtc.riken.go.jp) for further details.
FEATURES	
source	Location/Qualifiers 1..287 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="2210008P05" /clone_lib="Mus musculus stomach C57BL/6J adult" /sex="male" /tissue_type="stomach" /dev_stage="adult"
BASE COUNT	80 a 76 c 63 g 68 t
ORIGIN	
alignment_scores:	
Quality:	48.00 Length: 19
Ratio:	3.000 Gaps: 0

```

alignment_block:
US-08-653-294-31 x AV074366/rev ..
Align seg 1/1 to reverse of: AV074366 from: 1 to: 287

      2 ArgLeuAlaIleArgLeuAsnGluArgTyrArgLeuAlaIleArgLeuAs 18
      |||||:::|||||::: |||:::|||||::: ::|||:::
      89 AGACTGGGGATCAGCTCAGCTCAGGTTCCTCGCTCAGGCTCAGGCTTAG 40

      18 ngluArg 20
          :
          |||
      39 GCTCAGG 33

seq_name: gb_est22:AI070642

seq_documentation_block:
LOCUS      AI070642      345 bp      mRNA
DEFINITION UI-R-C2-mv-g-10-0-UI.s1 UI-R-C2 Rattus norvegicus cDNA clone
05-JUL-1999

```





ACCESSION AQ500194  
 VERSION AQ500194.1 GI:4705016  
 KEYWORDS GSS.  
 SOURCE baker's yeast.  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Saccharomycetales;  
 Saccharomycetaceae; Saccharomycetes.  
 REFERENCE 1 (bases 1 to 533)  
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,  
 desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,  
 Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,  
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.  
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
 Gene Disruption  
 JOURNAL Unpublished (1999)  
 COMMENT On Feb 19, 1999 this sequence version replaced gi:4143976.  
 Contact: Kumar A  
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
 Yale University  
 P.O. Box 208103, New Haven, CT 06520-8103, USA  
 Tel: 203 432 9949  
 Fax: 203 432 6161  
 Email: anuj.kumar@yale.edu  
 te of mtn-3xHA/lacZ insertion.  
 Seq primer: GGCCTCTCTCTTTTGAAGTAC  
 Class: transposon-tagged.  
 FEATURES Location/Qualifiers  
 source 1..533  
 /organism="Saccharomyces cerevisiae"  
 /db\_xref="taxon:4932"  
 /clone\_lib="mtn-3xHA/lacZ Insertion Library"  
 /lab\_host="E. coli"  
 /note="Vector: pHS56-Sal; A yeast genomic DNA library  
 (lacking mitochondrial DNA) was prepared in pHS56-Sal;  
 genomic DNA was size-fractionated (DNA of roughly 2-3 kb  
 in length) prior to cloning. This library was  
 subsequently mutagenized with a mtn-3xHA/lacZ  
 minitransposon containing lacZ, URA3, and tet resistance."  
 BASE COUNT 125 a 133 c 98 g 175 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 47.00 Length: 14  
 Ratio: 3.615 Gaps: 0  
 Percent Similarity: 92.857 Percent Identity: 64.286  
 alignment\_block:  
 US-08-653-294-31 x AQ500194 ..  
 Align seg 1/1 to: AQ500194 from: 1 to: 533  
 1 TtArgLeuAlaIleArgLeuAsnGluArgTyrArgLeuAla 14  
 |||||  
 193 TATAGTTGGCTCTAGCACTTTCACGGCAGTATCACTTGA 234  
 seq\_name: gb\_est41:AV266473  
 seq\_documentation\_block:  
 LOCUS AV266473 271 bp mRNA EST 05-NOV-1999  
 DEFINITION AV266473 RIKEN full-length enriched, adult male testis (DH10B) Mus  
 musculus cDNA clone 4930518A01 3', mRNA sequence.  
 ACCESSION AV266473  
 VERSION AV266473.1 GI:6254510  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 271)  
 AUTHORS Konno, H., Alzawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,  
 Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,  
 Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,  
 Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,  
 Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
 Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N.,  
 Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A.,  
 Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Konno, H., et al.)  
 Unpublished (1999)  
 On Jun 5, 1998 this sequence version replaced gi:3189489.  
 Contact: Yoshihide Hayashizaki  
 Genome Exploration Research Group, Life Science Tsukuba Center,  
 The Institute of Physical and Chemical Research (RIKEN), Genomic  
 Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: +81-298-36-9013  
 Fax: +81-298-36-9098  
 Email: genome-res@tc.riken.go.jp,  
 URL:http://genome.rtc.riken.go.jp/  
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
 Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and  
 Hayashizaki, Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
 Okazaki, Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (http://genome.rtc.riken.go.jp) for  
 further details.  
 FEATURES Location/Qualifiers  
 source 1..271  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RIKEN full-length enriched, adult male testis  
 (DH10B)"  
 /sex="male"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5'  
 GAGAGAGAGATTCGAGTTAAATAATCCCTCCCCCCCC 3']. cDNA  
 was cloned into the XhoI and BamHI sites. Vector: a  
 modified pBluescript KS(+) after bulk excision from Lambda  
 FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."  
 BASE COUNT 69 a 59 c 73 g 70 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 46.00 Length: 14  
 Ratio: 3.833 Gaps: 0  
 Percent Similarity: 85.714 Percent Identity: 57.143  
 alignment\_block:

```

US-08-653-294-31 x AV266473      ..
Align seg 1/1 to: AV266473 from: 1 to: 271
4 AlaIleArgLeuAsnGluAArgTyrArgLeuAlaIleArgLeu 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||
169 GCTGTAAAGTAAGTAATGCTAGGTATCGAGTTGCTGTACCCCTC 210
seq name: qb est37:AI998122

```

```

seq_name: gb_est37:AT998122

seq_documentation_block:
  LOCUS       330 bp                mRNA                08-SEP-1999
  DEFINITION   AT998122 A. thaliana, Columbia Col-0, rosette-1 Arabidopsis
               thaliana cDNA clone 701672303, mRNA sequence.
  ACCESSION   AT998122
  VERSION     AT998122.1 GI:5844957
  SOURCE      EST.
  ORGANISM    thale cress.
               Arabidopsis thaliana
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
               eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
               Arabidopsis.
  REFERENCE   1 (bases 1 to 330)
               Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
               Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
               Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D.,
               Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C.,
               Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S.,
               Nobriga, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and
               Hanson, D.
               Arabidopsis thaliana Gene Expression MicroArray
               Unpublished (1999)
               On Jun 5, 1998 this sequence version replaced gi:3189440.
  CONTACT     Contact: David Smoller, Ph.D.
               Genome Systems, Inc., a wholly owned subsidiary of Incyte
               Pharmaceuticals, Inc.
               4633 World Parkway Circle, St. Louis, MO 63134, USA
               Tel: 877-577-2733
               Fax: 314-427-3324
               Email: service@genomesystems.com.

```

FEATURES	source	BASE COUNT	ORIGIN
<p>1. 330</p> <p>location/Qualifiers</p> <p>/organism="Arabidopsis thaliana"</p> <p>/cultivar="Columbia Col-0"</p> <p>/db_xref="taxon:3702"</p> <p>/clone="701672303"</p> <p>/clone_lib="A. thaliana, Columbia Col-0, rosette-1"</p> <p>/tissue_type="rosette"</p> <p>/dev_stage="4 - 7 weeks"</p> <p>/note=vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."</p>	<p>80 a</p> <p>70 c</p> <p>68 g</p> <p>110 t</p> <p>2 others</p>		

```

alignment_scores:
  Quality: 46.00      Length: 19
  Ratio: 2.875       Gaps: 0
  Percent Similarity: 84.211  Percent Identity: 52.632

alignment_block:
  US-08-653-294-31 x AI998122      ..
Align seg 1/1 to: AI998122 from: 1 to: 330

```

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 04:05:44 ; Search time 133.56 seconds  
(without alignments)  
4.434 Million cell updates/sec

Title: US-08-653-294-32

Perfect score: 133

Sequence: 1 WDRETOICKAKAQTDRNLRIALRY 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	133	100.0	25	1	R41205	Peptide fragment o
2	133	100.0	25	1	R83090	HLA-B*2702 CTL modu
3	133	100.0	25	1	R95416	HLA-B*2702.60-84. C
4	133	100.0	25	1	R33794	Peptide B2702.60-8
5	118	88.7	25	1	R41221	Peptide fragment o
6	118	88.7	25	1	R83091	HLA-B*2702 CTL modu
7	118	88.7	25	1	R95417	HLA-B*2705.60-84. C
8	118	88.7	337	1	P70590	Sequence of the hu
9	118	88.7	362	1	P70155	Sequence encoded b
10	112	84.2	26	1	R83071	CTL modulating pep
11	109	82.0	21	1	W58992	Human HLA-B*27 alph
12	106	79.7	25	1	R48286	Peptide fragment o
13	106	79.7	25	1	R83093	HLA-B*38 CTL modul
14	106	79.7	25	1	R95422	HLA-B*38.6084. Comp
15	104	78.2	25	1	R95431	HLA-B*7.60-84. Comp
16	101	75.9	362	1	R03142	Sequence of HLA-Bw
17	95	71.4	184	1	Y06801	Peptide Seq ID No:
18	95	71.4	362	1	R03144	Sequence of HLA-B5
19	95	71.4	362	1	R12463	HLA-B*53 exon. HLA
20	88	66.2	274	1	P80911	Consensus sequence
21	74	55.6	15	1	R92912	HLA-B*2702 CTL modu
22	74	55.6	15	1	W33795	Peptide B2702.70-8
23	73	54.9	13	1	W58993	Human HLA-B*27 alph
24	72	54.1	25	1	R41207	Peptide fragment o
25	72	54.1	25	1	R83073	HLA-B*62 CTL modul
26	72	54.1	25	1	R95419	HLA-B*62.60-84. Co
27	70	52.6	25	1	R41206	Peptide fragment o
28	70	52.6	25	1	R83072	HLA-B*46 CTL modul
29	70	52.6	25	1	R95418	HLA-B*46.60-84. Co
30	70	52.6	45	1	R71629	HLA-alpha-1. Use o
31	70	52.6	366	1	R12465	HLA-C exon Cb-1. H
32	70	52.6	366	1	Y07033	Breast cancer asso
33	68	51.1	12	1	R41203	Peptide fragment w
34	66	49.6	362	1	R12464	HLA-B*35 antigen. H

35	65	48.9	366	1	R12466	HLA-C exon Cb-2. H
36	63	47.4	17	1	R71442	Human HLA-B*27-(62-
37	63	47.4	24	1	R71435	Human MHC I alpha
38	62	46.6	345	1	P83149	Probe F10-encoded
39	61	45.9	16	1	R50266	HLA B*27 hypervaria
40	60	45.1	25	1	R20116	MHC Class I-deriva
41	60	45.1	25	1	R69619	MHC-I peptide Dk-(
42	60	45.1	25	1	R71420	Human MHC I alpha
43	59	44.4	10	1	R41213	Peptide fragment o
44	59	44.4	10	1	R83076	HLA-B*2702/05 CTL m
45	59	44.4	10	1	R95424	HLA-B*2702/05.60-69

## ALIGNMENTS

### RESULT 1

ID R41205 standard; peptide; 25 AA.  
AC R41205;  
DT 15-MAR-1994 (first entry)  
DE Peptide fragment of Class I HLA peptide.  
KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
KW parasitic disease; cytotoxic T lymphocyte; modulation.  
OS Synthetic.  
PN W09317699-A.  
PD 16-SEP-1993. U01758.  
PF 25-FEB-1993; US-844716.  
PR 02-MAR-1992; US-844716.  
PI (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger CA, Krensky AM;  
DR WPI; 93-303134/38.  
PT New peptide(s) based on Class I HLA antigen domains - used for  
PT modulating cytotoxic T-lymphocyte activity towards targets  
PS Claim 8; Page 53; 61pp; English.  
CC The peptide (or a fragment of at least 10 amino acids, joined at at  
CC least one terminus to a sequence other than that of wild type HLA  
CC antigen) is used to modulate cytotoxic T-lymphocyte (CTL) activity,  
CC either by inhibition or stimulation. It can be used for  
CC inhibiting CTL toxicity in transplantations, for inducing CTL  
CC activity in parasitic diseases and neoplasia and in studies on viral  
CC infection. The peptide can also be used for identifying CTLs which  
CC bind to it and removing subsets of CTLs from a T-cell composition.  
SQ Sequence 25 AA;

Query Match 100.0%; Score 133; DB 1; Length 25;

Best Local Similarity 100.0%; Pred. No. 4.1e-14;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25

DB 1 WDRETOICKAKAQTDRNLRIALRY 25

### RESULT 2

ID R83090 standard; peptide; 25 AA.

AC R83090;

DT 16-MAY-1996 (first entry)

DE HLA-B\*2702 CTL modulating peptide (B2702.60-84).

KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;

KW immunosuppressant; graft versus host disorder; transplantation; therapy;

KW Class I MHC; HLA-B\*2702.

OS Synthetic.

PN W09526979-A1.

PD 12-OCT-1995.

PF 05-APR-1995; U04349.

PR 05-APR-1994; US-222851.

PI (STRD ) UNIV LELAND STANFORD JUNIOR.

PI Clayberger C, Krensky AM, Parham P;

DR WPI; 95-358582/46.

PT Extension of acceptance period of transplants from MHC unmatched

Query Match	100.0%;	Score 133;	DB 1;	Length 25;
Best Local Similarity	100.0%;	Pred. No. 4.1e-14;		
Matches 25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT	5
R41221	
ID	R41221 standard; peptide; 25 AA.
AC	R41221;
DT	15-MAR-1994 (first entry)
DE	Peptide fragment of HLA-B2705 antigen.
KW	Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;
KW	parasitic disease; cytotoxic T lymphocyte; modulation.
OS	Synthetic.
PN	WO9317699-A.
PD	16-SEP-1993.
PF	25-FEB-1993; U01758.
PR	02-MAR-1992; US-844716.
PA	(STRD ) UNIV LELAND STANFORD JUNIOR.



CC for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27  
 CC antibody in human serum, or to produce mono- or polyclonal HLA B27  
 CC antibodies for use in immunoassay.  
 SQ Sequence 337 AA;

Query Match 88.7%; Score 118; DB 1; Length 337;  
 Best Local Similarity 88.0%; Pred. No. 1.5e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 60 WDRETOICKAKAQTDRNLRIALRY 84

## RESULT 9

P70155 ID P70155 standard; protein: 362 AA.  
 AC P70155;  
 DT 10-MAR-1993 (revised)  
 DE Sequence encoded by genomic DNA encoding human histocompatibility  
 antigen HLA-B 27.  
 KW Ankylosing spondylitis; rheumatic disorder; diagnosis.  
 OS Homo sapiens.  
 PN EP-226069-A.  
 PD 24-JUN-1987.  
 PF 21-NOV-1986; 116139.  
 PR 01-JAN-1985; DE-542024.  
 PR 21-DEC-1985; DE-545576.  
 PA (BEHW) BEHRINGERWERKE AG.  
 PI Szots H, Weiss E, Dorner C, Lang M, Meo T, Riethmuller G;  
 DR WPI: 87-171469/25.  
 DR N-PSDB: N70225.  
 PT DNA coding for human histocompatibility antigen HLA-B 27 - useful  
 PT for diagnosis and antigen and antibody prodn.  
 PS Disclosure: p6; 13pp; German.  
 CC The DNA may be used to detect the HLA-B 27 gene (opt. mutated) in  
 CC human genetic material. The HLA-B 27 may be used to detect anti-HLA-  
 CC B 27 antibodies in human serum. The antibodies may be used to  
 CC determine HLA-B 27 levels in human serum, eg for diagnosis of  
 CC rheumatic disorders, esp. ankylosing spondylitis.  
 SQ Sequence 362 AA;

Query Match 88.7%; Score 118; DB 1; Length 362;  
 Best Local Similarity 88.0%; Pred. No. 1.6e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 84 WDRETOICKAKAQTDRNLRIALRY 108

## RESULT 10

R83071 ID R83071 standard; peptide: 26 AA.  
 AC R83071;  
 DT 16-MAY-1996 (first entry)  
 DE CTL modulating peptide #5.  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW class I MHC.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 18 /label= Asn, Asp  
 FT misc\_difference 21 /label= Ile, Thr  
 FT misc\_difference 22 /label= Ala, Leu  
 PN W0526979-A1.  
 PD 12-OCT-1995.

PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Farham P;  
 DR WPI: 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Disclosure: Page 9; 80pp; English.  
 CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. These sequences  
 CC can be used to extend the period of acceptance by a recipient of a  
 CC transplant from an MHC unmatched donor. The peptides are administered to  
 CC a patient in conjunction with a subtherapeutic amount of an  
 CC immunosuppressant. This is administered to the patient for a limited  
 CC period of time (compared to the lifetime administration for current  
 CC treatments). The peptides particularly modulate (or inhibit) the  
 CC activity of the cytotoxic T lymphocytes (CTLs) of the patient.  
 SQ Sequence 26 AA;

Query Match 84.2%; Score 112; DB 1; Length 26;  
 Best Local Similarity 84.0%; Pred. No. 7e-11;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 1 WDRETOICKAKAQTDRNLRIALRY 25

## RESULT 11

W58992 ID W58992 standard; peptide: 21 AA.  
 AC W58992;  
 DT 30-JUL-1998 (first entry)  
 DE Human HLA-B27 alpha-1 domain (aa. 60-80) peptide B27PA.  
 DE Human leucocyte antigen; HLA; alpha-1 domain; alpha-2 domain; HLA-B27;  
 KW keratin VI; diagnosis; treatment; HLA-dependent autoimmune disease;  
 KW Class I-associated autoimmune disease; Iritis; uveitis; psoriasis;  
 KW Class II-associated autoimmune disease; Iritis; uveitis; psoriasis;  
 KW ankylosing spondylitis; antigen; regulator.  
 OS Homo sapiens.  
 PN W09812221-A1.  
 PD 26-MAR-1998.  
 PF 18-SEP-1997; E05124.  
 PR 18-SEP-1996; DE-038108.  
 PA (WILD/) WILDNER G.  
 PI Wildner G;  
 DR WPI: 98-217206/19.  
 PT Treatment and diagnosis of HLA-dependent autoimmune disease - using  
 PT peptide(s) derived from HLA-B27 or human keratin VI, e.g. for  
 PT Iritis, uveitis, rheumatic disease etc.  
 PS Claim 2; Page 9; 57pp; German.  
 CC W58992-W59001 are peptides isolated from the alpha-1 or alpha-2 domain  
 CC of HLA-B27 or human keratin IV and are used in methods for diagnosis  
 CC and/or treatment of HLA (human leucocyte antigen)-dependent autoimmune  
 CC diseases. Such peptides have implications in the treatment of cases  
 CC of Class I and II-associated autoimmune diseases, especially non-HLA-B27  
 CC diseases (Iritis, uveitis, psoriasis and rheumatic diseases such as  
 CC arthritis, psoriatic arthritis or juvenile rheumatoid arthritis) and  
 CC HLA-B27 diseases e.g. ankylosing spondylitis (AS). These fragments can  
 CC be administered intravenously, subcutaneously or intramuscularly, or to  
 CC the mucosa (orally, or as nasal or pulmonary spray). This method results  
 CC in antigens that are more specific than complete protein antigens, so  
 CC provide a more exact classification of disease, and thus a more specific  
 CC treatment. In the case of keratin-derived peptides, they also have better  
 CC solubility. Compared with complete proteins, peptides are easier to  
 CC prepare, do not require recovery from natural tissue or recombinant  
 CC methods of production, have better storage stability when dry, are less  
 CC likely to cause allergy and have a regulatory effect on the immune  
 CC system.  
 SQ Sequence 21 AA;

RESULT	12
R48286	
ID	R48286 standard; peptide; 25 AA.
AC	R48286;
DT	15-MAR-1994 (first entry)
DE	Peptide fragment of HLA-B38 antigen.
KW	Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;

OS	Synthetic.
PN	W09317699-A.
PD	16-SEP-1993.
PF	25-FEB-1993.
PR	02-MAR-1992.
PA	(STRD ) UNIV LELAND STANFORD JUNIOR.
PI	Clayberger CA, Krensky AM;
DR	WPI; 93-303134/38.
PPT	New peptide(s) based on Class I HLA antigen domains - used for
PPT	modulating cytotoxic T-lymphocyte activity towards targets
PS	Example 13; Page 39; 61pp; English.
CCC	The peptide is used to modulate cytotoxic T-lymphocyte (CTL)
CCC	activity, either by inhibition or stimulation. It can be used for
CCC	inhibiting CTL toxicity in transplantations, for inducing CTL
CCC	activity in parasitic diseases and neoplasia and in studies on viral
CCC	infection. The peptide can also be used for identifying CRIs which
CCC	bind to it and removing subsets of CTLs from a T-cell composition.
CCC	This peptide is derived from the HLA-B38 antigen and corresponds
CCC	to the amino acid positions 60-84 of that antigen.
SQ	Sequence 25 AA;

```

Query Match          79.7%; Score 106; DB 1; Length 25;
Best Local Similarity 80.0%; Pred. No. 5.6e-10;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 WDRETOICKAKAQTDRENLRALRY 25
          ||| ||||| || ||||| |||
Ddb      1 WDRETOICKTNTQTYRENLRALRY 25

RESULT 13
R83093 ID R83093 standard; peptide: 25 AA.
R83093 AC R83093;
DT 16-MAY-1996 (first entry)
DE HLAB38 CTL modulating peptide (B38 6084).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLAB38.
KW OS Synthetic.
PN WO9526979-A1.
PD 12-OCT-1995.

```

05-APR-1994; US-222851.  
(STRD) UNIV LELAND STANFORD JUNIOR.  
PA Clayberger C, Krensky AN, Patham P;  
DR WPI: 95-358582/46.  
PT Extension of acceptance period of transplants from MHC unmatched  
PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
PT host  
Example 13: Page 32: 80pp: English.  
PS R83061-R83085, R83090-R83096 and R82907-R82913 represent fragments of  
CC class I major histocompatibility complex (MHC) antigens. This sequence  
CC corresponds to residues 60-84 of the alpha-1 domain of the class I MHC  
CC

R95431  
 ID R95431 standard; peptide; 25 AA.  
 AC R95431;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B\*7.60-84.  
 KW HLA: p74: alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN W09513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Compsns. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 12; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B\*7.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs) by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 78.28; Score 104; DB 1; Length 25;  
 Best Local Similarity 95.08; Pred. No. 1.1e-09;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 WDRETIQCKAKAQTDRENLR 20  
 |||||  
 Db 1 WDRETIQCKAKAQTDRESLR 20  
 |||||

Search completed: February 8, 2000, 04:05:44  
 Job time: 9361 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 18:04:40 ; Search time 111.22 Seconds  
(without alignments)  
10.603 Million cell updates/sec

Title: US-08-653-294-32  
Perfect score: 133  
Sequence: 1 WDRETQICKAKAQTDRENRLRY 25

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

PIR\_62:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	93.2	354	2 I59308	class I histocompa
2	124	93.2	354	2 I80167	class I histocompa
3	120	90.2	359	1 HLRU12	MHC class I histoc
4	118	88.7	338	2 I56116	MHC HLA-B27-HS - h
5	118	88.7	362	1 HLRU32	MHC class I histoc
6	118	88.7	362	2 I37485	human lymphocyte a
7	118	88.7	362	2 I37515	MHC class I histoc
8	118	88.7	362	2 I54289	MHC HLA-B27d - hum
9	114	85.7	362	2 JH0541	class I histocompa
10	114	85.7	362	2 JH0539	class I histocompa
11	114	85.7	362	2 JH0540	class I histocompa
12	107	80.5	355	2 I80169	class I histocompa
13	107	80.5	355	2 I80171	class I histocompa
14	107	80.5	363	2 S07113	class I histocompa
15	107	80.5	363	2 S03537	class I histocompa
16	106	79.7	274	2 I54463	MHC HLA-B38 chain
17	101	75.9	362	2 B30345	MHC class I histoc
18	101	75.9	362	2 S24434	class I histocompa
19	98	73.7	354	2 I80168	class I histocompa
20	98	73.7	362	2 I62045	gene HLA B-1517 pr
21	96	72.2	355	2 I37516	HLA-B alpha-chain
22	96	72.2	362	2 S25415	class I histocompa
23	96	72.2	362	2 A45850	MHC class I histoc
24	96	72.2	362	2 I61861	MHC HLA-B44.2 chai
25	96	72.2	362	2 I84486	transmembrane glyco
26	96	72.2	362	2 I54442	MHC class I histoc
27	95	71.4	273	2 I38509	MHC class I histoc
28	95	71.4	362	2 A45834	MHC class I histoc
29	95	71.4	362	2 A30345	MHC class I histoc
30	95	71.4	362	2 I59633	MHC HLA-B transmem

31	95	71.4	362	2 I37120	MHC class I histoc
32	95	71.4	365	2 S77963	MHC class I histoc
33	95	71.4	365	2 I34416	HLA-AW24 protein -
34	95	71.4	365	2 I34493	MHC class I histoc
35	94	70.7	274	1 HLRU32	MHC class I histoc
36	93	69.9	364	2 A35997	MHC class I histoc
37	92	69.2	137	2 I80174	class I histocompa
38	91	68.4	359	1 HLRU34	MHC class I histoc
39	91	68.4	362	1 HLRU38	MHC class I histoc
40	91	68.4	362	2 I84490	lymphocyte antigen
41	91	68.4	362	2 I37521	HLA-Bw57.2 antigen
42	91	68.4	363	1 S42102	MHC class I histoc
43	91	68.4	364	2 D35997	MHC class I histoc
44	90	67.7	354	2 I80165	class I histocompa
45	89	66.9	365	2 JH0537	class I histocompa

## ALIGNMENTS

RESULT 1  
I59308

class I histocompatibility antigen - pygmy chimpanzee (fragment)

C:Species: Pan paniscus (pygmy chimpanzee, bonobo)

C>Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 23-Jul-1999

C:Accession: I59308

R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat

Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994

A:Title: A uniquely high level of recombination at the HLA-B locus.

A:Reference number: I59308; MUID:94286544

A:Accession: I59308

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-354 <RES>

A:Cross-references: EMBL:U05575; NID:9454767; PIDN:AAA50178.1; PID:9454768

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 93.2%; Score 124; DB 2; Length 354;  
Best Local Similarity 92.0%; Pred. No. 9.4e-12;  
Matches 23; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WDRETQICKAKAQTDRENRLRY 25

Db 76 WDRTQICKAKAQTDRENRLRY 100

RESULT 2

class I histocompatibility antigen - pygmy chimpanzee (fragment)

C:Species: Pan paniscus (pygmy chimpanzee, bonobo)

C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 23-Jul-1999

C:Accession: I80167

R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat

Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994

A:Title: A uniquely high level of recombination at the HLA-B locus.

A:Reference number: I59308; MUID:94286544

A:Accession: I80167

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-354 <RES>

A:Cross-references: EMBL:U05578; NID:9454773; PIDN:AAA50181.1; PID:9454774

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 93.2%; Score 124; DB 2; Length 354;  
Best Local Similarity 92.0%; Pred. No. 9.4e-12;  
Matches 23; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WDRETQICKAKAQTDRENRLRY 25

Db 76 WDRTQICKAKAQTDRENRLRY 100

## RESULT 3

HLH012

MHC class I histocompatibility antigen HLA alpha chain precursor (clone pHLA 12.4) - hum

C:Species: Homo sapiens (man)

C:Date: 05-Apr-1983 #sequence\_revision 05-Apr-1983 #text\_change 22-Jun-1999

C:Accession: A02189

R:Malissen, M.; Malissen, B.; Jordan, B.R.

Proc. Natl. Acad. Sci. U.S.A. 79, 893-897, 1982

A:Title: Exon/intron organization and complete nucleotide sequence of an HLA gene.

A:Reference number: A02189; MUID:82151002

A:Accession: A02189

A:Molecule type: DNA

A:Residues: 1-359 &lt;MAL&gt;

A:Cross-references: GB:J00191; GB:V00526; NID:g187600; PIDN:AAA36218.1; PID:g386873

C:Comment: The seven exons correspond approximately to the domain structure of this chain

C:Genetics:

A:Map position: 6p21.3

A:Introns: 22/1: 112/1: 204/1: 296/1: 335/1: 346/1

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterodimer; transmembrane protein; transplanted

F:1-21/Domain: signal sequence #status predicted &lt;SIG&gt;

F:22-359/Product: class I histocompatibility antigen HLA alpha chain #status predicted &lt;

F:22-304/Domain: extracellular #status predicted &lt;EXT&gt;

F:22-111/Domain: alpha-1 &lt;EX1&gt;

F:112-203/Domain: alpha-2 &lt;EX2&gt;

F:217-282/Domain: immunoglobulin homology &lt;IMM&gt;

F:305-329/Domain: transmembrane #status predicted &lt;TM&gt;

F:335-359/Domain: intracellular #status predicted &lt;INT&gt;

F:107/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:224-280/Disulfide bonds: #status predicted

Query Match 90.2%; Score 120; DB 1; Length 359;

Best Local Similarity 88.0%; Pred. No. 4e-11;

Matches 22; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 WDRETOICKAKAQTDERNLRIALRY 25

Db 81 WDRETOICKAKAQTDERNLRIALRY 105

## RESULT 4

I56116

MHC HLA-B27-HS - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999

C:Accession: I56116

R:Choo, S.Y.; Fan, L.A.; Hansen, J.A.

J. Immunol. 147, 174-180, 1991

A:Title: A novel HLA-B27 allele maps B27 allospecificity to the region around position 7

A:Reference number: I56116; MUID:91268545

A:Accession: I56116

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-338 &lt;RES&gt;

A:Cross-references: GB:M62852; NID:g187760; PIDN:AAA59647.1; PID:g187761

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

## Query Match

Best Local Similarity 88.7%; Score 118; DB 2; Length 338;

Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 WDRETOICKAKAQTDERNLRIALRY 25

Db 60 WDRETOICKAKAQTDERNLRIALRY 84

## RESULT 5

HLH082

MHC class I histocompatibility antigen HLA-B27 alpha chain precursor - human

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999

C:Accession: I56116

R:Choo, S.Y.; Fan, L.A.; Hansen, J.A.

J. Immunol. 147, 174-180, 1991

A:Title: A novel HLA-B27 allele maps B27 allospecificity to the region around position 7

A:Reference number: I56116; MUID:91268545

A:Accession: I56116

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-338 &lt;RES&gt;

A:Cross-references: GB:M62852; NID:g187760; PIDN:AAA59647.1; PID:g187761

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

## Query Match

Best Local Similarity 88.0%; Score 118; DB 2; Length 338;

Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

C:Date: 13-Aug-1986 #sequence\_revision 28-Apr-1995 #text\_change 22-Jun-1999  
C:Accession: S07441; B25092; B25092; A94087; S4180; S44942; A90493; B24741; I55965;  
R:Weiss, E.H.; Kuon, W.; Doerner, C.; Lang, M.; Riethmuller, G.  
Immunobiology 170, 367-380, 1985

A:Title: Organization, sequence and expression of the HLA-B27 gene: a molecular approach  
A:Reference number: S07441; MUID:86138405  
A:Accession: S07441

A:Molecule type: DNA

A:Residues: 1-362 &lt;WEI&gt;

A:Cross-references: EMBL:X03945

A:Note: The authors translated the codon GAC for residue 61 as Ala and the codon CAG  
A:Note: this allele is designated B\*27052 (formerly 27W)  
R:Seemann, G.H.A.; Rein, R.S.; Brown, C.S.; Ploegh, H.L.  
EMBO J. 5, 547-552, 1986

A:Title: Gene conversion-like mechanisms may generate polymorphism in human class I g  
A:Reference number: A91061; MUID:86220133  
A:Accession: A25092

A:Molecule type: DNA

A:Residues: 1-362 &lt;SEE&gt;

A:Cross-references: GB:X03665; NID:g22250; PIDN:CAA27302.1; PID:g871297

A:Note: this allele is designated B\*27051 (formerly 27W)

A:Accession: B25092

A:Molecule type: DNA

A:Residues: 1-100, 'N', '102-103, 'IA', '106-362 &lt;SE2&gt;

A:Cross-references: GB:X03664; NID:g32236; PIDN:CAA27301.1; PID:g871296

A:Note: this allele is designated B\*2702 (formerly 27K)

R:Stoets, H.; Riethmuller, G.; Weiss, E.; Meo, T.  
Proc. Natl. Acad. Sci. U.S.A. 83, 1428-1432, 1986

A:Title: Complete sequence of HLA-B27 cDNA identified through the characterization of  
A:Reference number: A94087; MUID:86149317  
A:Accession: A94087

A:Molecule type: mRNA

A:Residues: 25-205, 'V', '207-362 &lt;SZO&gt;

A:Cross-references: GB:M12678

A:Note: this allele is designated B\*27052 (formerly 27W)

R:Viaches, C.

submitted to the EMBL Data Library, June 1993

A:Reference number: S34180

A:Accession: S34180

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-100, 'S', '102-137, 'D', '139, 'Y', '141-175, 'E', '177-362 &lt;VIL&gt;

A:Note: this allele is designated B\*2706

R:D'Amato, M.; Sorrentino, R.

submitted to the EMBL Data Library, May 1994

A:Description: Identification of a novel HLA-B27 subtype by restriction analysis of a  
A:Reference number: S44942  
A:Accession: S44942

A:Molecule type: mRNA

A:Residues: 1-139, 'H', '141-362 &lt;DAM&gt;

A:Cross-references: EMBL:Z33453; NID:g486652; PIDN:CAA83876.1; PID:g486653

R:Esquerre, A.; Bragado, R.; Vega, M.A.; Strominger, J.L.; Woody, J.; Lopez de Castro  
Biochemistry 24, 1733-1741, 1985

A:Title: Primary structure of papain-solubilized human histocompatibility antigen HLA  
A:Reference number: A90493; MUID:85226361  
A:Accession: A90493

A:Molecule type: protein

A:Residues: 25-265, 'E', '267-295 &lt;EZQ&gt;

R:Vega, M.A.; Esquerre, A.; Rojo, S.; Aparicio, P.; Bragado, R.; Lopez de Castro, J.A.

Proc. Natl. Acad. Sci. U.S.A. 82, 7394-7398, 1985

A:Title: Structural analysis of an HLA-B27 functional variant: identification of resi  
A:Reference number: A94070; MUID:86042671  
A:Accession: B24741

A:Molecule type: protein

A:Residues: 86-100, 'N', '102-103, 'IA', '106-107; 171-181 &lt;VEG&gt;

R:Coppin, H.L.; McDevitt, H.O.

J. Immunol. 137, 2168-2172, 1986

A:Title: Absence of polymorphism between HLA-B27 genomic exon sequences isolated from  
A:Reference number: I55965; MUID:87009855  
A:Accession: I55965

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA



A:Residues: 25-298 <RES>  
A:Cross-references: GB:M14013; NID:g187743; PIDN:AAA59643.1; PID:g187744  
R:Blaszzyk, R.; Weber, M.; Salama, A.  
Submitted to the EMBL Data Library, January 1995  
A:Reference number: S52291  
A:Accession: S52291  
A:Molecule type: DNA  
A:Residues: 116-192 <BLA>  
A:Cross-references: EMBL:X83737  
C:Comment: This allele for HLA-B correlates with the development of ankylosing spondylitis  
C:Genetics:  
A:Gene: GDB:HLA-B  
A:Cross-references: GDB:120048; OMIM:142830  
A:Map position: 6p21.3-6p21.3  
A:Introns: 25/1; 114/1; 207/1; 299/1; 338/1; 349/1  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: ankylosing spondylitis; duplication; glycoprotein; heterodimer; transmembran  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen HLA-B27 alpha chain #status predic  
F:25-307/Domain: extracellular #status predicted <EXT>  
F:25-114/Domain: alpha-1 <EX1>  
F:115-206/Domain: alpha-2 <EX2>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:308-331/Domain: transmembrane #status predicted <TMN>  
F:332-362/Domain: intracellular #status predicted <INT>  
F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:125-188,227-283/Disulfide bonds: #status experimental

Query Match 88.7%; Score 118; DB 1; Length 362;  
Best Local Similarity 88.0%; Pred. No. 8.2e-11;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLALRY 25  
|||||  
Db 84 WDRETOICKAKAQTDRDLRLTRY 108

RESULT 6  
I37485  
human lymphocyte antigen HLA-B27 - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999  
C:Accession: I37485  
R:Del Porto, P.; D'Amato, M.; Fiorillo, M.T.; Tuosto, L.; Piccolella, E.; Sorrentino, R.  
J. Immunol. 153, 3093-3100, 1994  
A:Title: Identification of a novel HLA-B27 subtype by restriction analysis of a cytotoxi  
A:Reference number: I37485; MUID:94375872  
A:Accession: I37485  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-362 <RES>  
A:Cross-references: EMBL:Z33453; NID:g486652; PIDN:CAA83876.1; PID:g486653  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 88.7%; Score 118; DB 2; Length 362;  
Best Local Similarity 88.0%; Pred. No. 8.2e-11;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLALRY 25  
|||||  
Db 84 WDRETOICKAKAQTDRDLRLTRY 108

RESULT 7  
I37515  
MHC class I histocompatibility antigen HLA-B\*2706 alpha chain precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 04-Oct-1996 #sequence\_revision 04-Oct-1996 #text\_change 23-Jul-1999  
C:Accession: I37515  
R:Vilches, C.; de Pablo, R.; Kreisler, M.  
Immunogenetics 39, 219, 1994

A:Title: Nucleotide sequence of HLA-B\*2706.  
A:Reference number: I37515; MUID:94102824  
A:Accession: I37515  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-362 <RES>  
A:Cross-references: EMBL:X73578; NID:g975658; PIDN:CAA51980.1; PID:g975659  
C:Genetics:  
A:Gene: GDB:HLA-B  
A:Cross-references: GDB:120048; OMIM:142830  
A:Map position: 6p21.3-6p21.3  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 88.7%; Score 118; DB 2; Length 362;  
Best Local Similarity 88.0%; Pred. No. 8.2e-11;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLALRY 25  
|||||  
Db 84 WDRETOICKAKAQTDRDLRLTRY 108

RESULT 8  
I54289  
MHC HLA-B27d - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999  
C:Accession: I54289  
R:Choo, S.Y.; St. John, T.; Orr, H.T.; Hansen, J.A.  
Hum. Immunol. 21, 209-219, 1988  
A:Title: Molecular analysis of the variant alloantigen HLA-B27d (HLA-B\*2703) identifi  
A:Reference number: I54289; MUID:98227491  
A:Accession: I54289  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-362 <RES>  
A:Cross-references: GB:M54883; NID:g187663; PIDN:AAA59616.1; PID:g187664  
C:Genetics:  
A:Gene: GDB:HLA-B  
A:Cross-references: GDB:120048; OMIM:142830  
A:Map position: 6p21.3-6p21.3  
A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 88.7%; Score 118; DB 2; Length 362;  
Best Local Similarity 88.0%; Pred. No. 8.2e-11;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLALRY 25  
|||||  
Db 84 WDRETOICKAKAQTDRDLRLTRY 108

RESULT 9  
JH0541  
Class I histocompatibility antigen Gogo-B0103 heavy chain precursor - lowland gorilla  
C:Species: Gorilla gorilla gorilla (lowland gorilla)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999  
C:Accession: JH0541  
R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.  
J. Exp. Med. 174, 1491-1509, 1991  
A:Title: Gorilla class I major histocompatibility complex alleles: comparison to huma  
A:Reference number: JH0534; MUID:92078860  
A:Accession: JH0541  
A:Molecule type: DNA  
A:Residues: 1-362 <LAW>  
A:Cross-references: EMBL:X60254; NID:g22869; PIDN:CAA42806.1; PID:g22870  
A:Experimental source: EBV-transformed B cell  
C:Genetics:  
A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C;Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-B0103 #status pre  
F:25-114/Domain: alpha-1 <AL1>  
F:115-206/Domain: alpha-2 <AL2>  
F:207-298/Domain: alpha-3 <AL3>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-362/Domain: intracellular #status predicted <INT>

Query Match 85.7%; Score 114; DB 2; Length 362;  
Best Local Similarity 88.0%; Pred. No. 3.5e-10;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
||||| ||:|||||  
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 10  
JH0539  
class I histocompatibility antigen Gogo-B0101 heavy chain precursor - lowland gorilla  
C:Species: Gorilla gorilla gorilla (lowland gorilla)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999  
C:Accession: JH0539  
R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.  
J. Exp. Med. 174, 1491-1509, 1991  
A:Title: Gorilla class I major histocompatibility complex alleles: comparison to human a  
A:Reference number: JH0534; MUID:92078860  
A:Accession: JH0539  
A:Molecule type: DNA  
A:Residues: 1-362 <LAW>  
A:Cross-references: EMBL:X60525; NID:g22865; PIDN:CAA42807.1; PID:g22866  
A:Experimental source: EBV-transformed B cell  
C:Genetics: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-B0101 #status pre  
F:25-114/Domain: alpha-1 <AL1>  
F:115-206/Domain: alpha-2 <AL2>  
F:207-298/Domain: alpha-3 <AL3>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-362/Domain: intracellular #status predicted <INT>

Query Match 85.7%; Score 114; DB 2; Length 362;  
Best Local Similarity 88.0%; Pred. No. 3.5e-10;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
||||| ||:|||||  
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 11  
JH0540  
class I histocompatibility antigen Gogo-B0102 heavy chain precursor - lowland gorilla  
C:Species: Gorilla gorilla gorilla (lowland gorilla)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999  
C:Accession: JH0540  
R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.  
J. Exp. Med. 174, 1491-1509, 1991  
A:Title: Gorilla class I major histocompatibility complex alleles: comparison to human a  
A:Reference number: JH0534; MUID:92078860  
A:Accession: JH0540  
A:Molecule type: DNA  
A:Residues: 1-362 <LAW>  
A:Cross-references: EMBL:X60593; NID:g22867; PIDN:CAA43101.1; PID:g22868  
A:Experimental source: EBV-transformed B cell  
C:Genetics: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1

C;Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C;Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-B0102 #status  
F:25-114/Domain: alpha-1 <AL1>  
F:115-206/Domain: alpha-2 <AL2>  
F:207-298/Domain: alpha-3 <AL3>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-362/Domain: intracellular #status predicted <INT>

Query Match 85.7%; Score 114; DB 2; Length 362;  
Best Local Similarity 88.0%; Pred. No. 3.5e-10;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
||||| ||:|||||  
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 12  
I80169  
class I histocompatibility antigen - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 23-Jul-1999  
C:Accession: I80169  
R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat  
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994  
A:Title: A uniquely high level of recombination at the HLA-B locus.  
A:Reference number: 159308; MUID:94286544  
A:Accession: I80169  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-355 <RES>  
A:Cross-references: EMBL:U05580; NID:g45477; PIDN:AAA50183.1; PID:g454778  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 80.5%; Score 107; DB 2; Length 355;  
Best Local Similarity 84.0%; Pred. No. 4.1e-09;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
||||| ||:|||||  
Db 76 WDRETRNMKASAQTDRNLRIALRY 100

RESULT 13  
I80171  
class I histocompatibility antigen - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 23-Jul-1999  
C:Accession: I80171  
R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat  
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994  
A:Title: A uniquely high level of recombination at the HLA-B locus.  
A:Reference number: 159308; MUID:94286544  
A:Accession: I80171  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-355 <RES>  
A:Cross-references: EMBL:U05582; NID:g454781; PIDN:AAA50185.1; PID:g454782  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 80.5%; Score 107; DB 2; Length 355;  
Best Local Similarity 84.0%; Pred. No. 4.1e-09;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
||||| ||:|||||  
Db 76 WDRETRNMKASAQTDRNLRIALRY 100

## RESULT 14

S07113  
Class I histocompatibility antigen Ch39 alpha chain precursor - chimpanzee  
N:Alternate names: MHC ChLA chain  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999  
C:Accession: S07113; I36957  
R:Lawlor, D.A.; Ward, F.E.; Ennis, P.D.; Jackson, A.P.; Parham, P.  
Nature 335, 268-271, 1988  
A:Title: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees.  
A:Reference number: S06424; MUID:88319000  
A:Accession: S07113  
A:Molecule type: mRNA  
A:Residues: 1-363 <LAW>  
R:Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.  
J. Immunol. 142, 3937-3950, 1989  
A:Title: Diversity and diversification of HLA-A,B,C alleles.  
A:Reference number: I36956; MUID:89235215  
A:Accession: I36957  
A:Molecule type: mRNA  
A:Residues: 1-363 <RES>  
A:Cross-references: GB:M24045; NID:g176814; PIDN:AAA35424.1; PID:g176815  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; membrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-363/Product: class I histocompatibility antigen Ch39 alpha chain #status predicted  
F:25-114/Domain: alpha-1 #status predicted <EX1>  
F:115-206/Domain: alpha-2 #status predicted <EX2>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:307-332/Domain: transmembrane #status predicted <TMM>  
F:333-363/Domain: intracellular #status predicted <INT>  
F:110/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:125-188,227-283/Disulfide bonds: #status predicted

Query Match 80.5%; Score 107; DB 2; Length 363;

Best Local Similarity 84.0%; Pred. No. 4.2e-09;

Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25

|||||: || |||||

Db 84 WDRETRNVKASQTDRENLRIALRY 108

## RESULT 15

S03537  
Class I histocompatibility antigen ChLA-B1 alpha chain precursor - chimpanzee  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 18-Oct-1996  
C:Accession: S03537  
R:Mayer, W.E.; Jonker, M.; Klein, D.; Ivanyi, P.; van Seventer, G.; Klein, J.  
EMBO J. 7, 2765-2774, 1988  
A:Title: Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species polymorphism.  
A:Reference number: S01171; MUID:89030641  
A:Accession: S03537  
A:Molecule type: mRNA  
A:Residues: 1-363 <MAY>  
A:Cross-references: EMBL:X13115  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-363/Product: class I histocompatibility antigen alpha chain #status predicted <MAT>  
F:25-298/Domain: extracellular #status predicted <EXT>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-338/Domain: transmembrane #status predicted <TMM>  
F:339-363/Domain: intracellular #status predicted <INT>

Query Match 80.5%; Score 107; DB 2; Length 363;

Best Local Similarity 84.0%; Pred. No. 4.2e-09;

Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
|||||: || |||||

Db 84 WDRETRNVKASQTDRENLRIALRY 108

Search completed: February 7, 2000, 18:04:41  
Job time: 22207 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:26:01 ; Search time 68.63 Seconds  
(without alignments)  
10.879 Million cell updates/sec

Title: US-08-653-294-32

Perfect score: 133  
Sequence: 1 WDRETIQCKAKAQTDRNLRIALRY 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	133	100.0	362	1	1B15_HUMAN
2	120	90.2	362	1	HLAH_HUMAN
3	118	88.7	338	1	1B20_HUMAN
4	118	88.7	361	1	1B14_HUMAN
5	118	88.7	362	1	1B16_HUMAN
6	118	88.7	362	1	1B18_HUMAN
7	118	88.7	362	1	1B19_HUMAN
8	114	85.7	362	1	1B01_GORGO
9	114	85.7	362	1	1B02_GORGO
10	114	85.7	362	1	1B03_GORGO
11	107	80.5	359	1	1B01_PANTR
12	101	75.9	362	1	1B47_HUMAN
13	101	75.9	362	1	1B53_HUMAN
14	96	72.2	362	1	1B05_HUMAN
15	96	72.2	362	1	1B41_HUMAN
16	96	72.2	362	1	1B42_HUMAN
17	95	71.4	362	1	1B49_HUMAN
18	95	71.4	362	1	1B52_HUMAN
19	95	71.4	362	1	1B54_HUMAN
20	95	71.4	365	1	1A23_HUMAN
21	95	71.4	365	1	1A24_HUMAN
22	94	70.7	365	1	1A32_HUMAN
23	93	69.9	365	1	1A25_HUMAN
24	91	68.4	359	1	1B40_HUMAN
25	91	68.4	362	1	1B60_HUMAN
26	91	68.4	362	1	1B61_HUMAN
27	91	68.4	362	1	1B62_HUMAN
28	89	66.9	365	1	1A04_GORGO
29	86	64.7	362	1	1B07_HUMAN
30	86	64.7	362	1	1B08_HUMAN
31	86	64.7	362	1	1B29_HUMAN
32	86	64.7	362	1	1B31_HUMAN
33	86	64.7	362	1	1B45_HUMAN
34	84	63.2	362	1	1B02_HUMAN

```

35      84 63.2 362 1 1B39_HUMAN      P30480 homo sapien
36      84 63.2 362 1 1B55_HUMAN      P30492 homo sapien
37      84 63.2 362 1 1B56_HUMAN      P30493 homo sapien
38      84 63.2 362 1 1B57_HUMAN      P30494 homo sapien
39      84 63.2 362 1 1B58_HUMAN      P30495 homo sapien
40      84 63.2 362 1 1B59_HUMAN      P30496 homo sapien
41      81 60.9 362 1 1B32_HUMAN      P30476 homo sapien
42      79 59.4 361 1 H1A1A_RABIT      P01894 oryctolagus
43      78 59.4 361 1 H1A1B_RABIT      P06140 oryctolagus
44      76 57.1 362 1 1A02_PANTR      P16210 pan troglod
45      76 57.1 362 1 1B02_PANTR      P13751 pan troglod

```

## ALIGNMENTS

```

RESULT 1
ID 1B15_HUMAN STANDARD; PRT; 362 AA.
AC P10317;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2702 ALPHA CHAIN
DE PRECURSOR (B*27K) (B27.2).
GN HLA-B OR HLAB
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86220133.
RA SEEMANN G.H.A., REIN R.S., BROWN C.S., PLOEGH H.L.;
RT "Gene conversion-like mechanisms may generate polymorphism in human
RT class I genes."
RL EMBO J. 5:547-552(1986).
RN [2]
RP SEQUENCE FROM N.A.
RA PARHAM P., ARNETT K.L., ADAMS E.J.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RX SEQUENCE OF 86-107 AND 171-181.
RX MEDLINE; 86042671.
RA VEGA M.A., EQUERRA A., ROJO S., APARICIO P., BRAGADO R.,
RA LOPEZ DE CASTRO J.A.;
RT "Structural analysis of an HLA-B27 functional variant: identification
RT of residues that contribute to the specificity of recognition by
RT cytolytic T lymphocytes."
RL Proc. Natl. Acad. Sci. U.S.A. 82:7394-7398(1985).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROCROBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X03664; CAA27301.1;
CC EMBL; X03667; CAA27301.1; JOINED.
CC EMBL; L38504; AAA69724.1;
CC PIR; B25092; HLHUBK.
CC HSSP; P03989; LHSA.
CC MIM; 142830;
CC PROSITE; PS00290; IG_MHC; 1.
CC PFAM; PF00047; Ig; 1.
CC PFAM; PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 24

```

FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 25 114 B-27 B\*2702 ALPHA CHAIN.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40397 MW; 9798FOBB CRC32;

Query Match 100.0%; Score 133; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 DB 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 2  
 ID HLA\_HUMAN STANDARD; PRT; 362 AA.  
 AC P01893;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-DEC-1999 (Rel. 39, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN H PRECURSOR  
 DE (HLA-AR) (HLA-12.4).  
 GN HLA-H OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 82151002.  
 RA MALISEN M., MALISEN B., JORDAN B.R.;  
 RT "Exon/Intron organization and complete nucleotide sequence of an HLA  
 gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:893-897(1982).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM. COULD BE THE PRODUCT OF A PSEUDOGENE.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: J00191; AAA36218.1; ALT\_INIT.  
 DR PIR: A02189; HLH012.  
 DR HSP: P03989; IHSA.  
 DR MIN: 142800; -.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC.I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 25 114 ALPHA CHAIN H.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.

FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40850 MW; 5E610F63 CRC32;  
 Query Match 90.2%; Score 120; DB 1; Length 362;  
 Best Local Similarity 88.0%; Pred. No. 4.8e-11;  
 Matches 22; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 DB 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 3  
 ID B20\_HUMAN STANDARD; PRT; 338 AA.  
 AC P30467;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B\*2707 ALPHA CHAIN  
 DE (B27-HS).  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91268545.  
 RA CHOO Y.S., FAN L.A., HANSEN J.A.;  
 RT "A novel HLA-B27 allele maps B27 allospecificity to the region around  
 position 70 in the alpha 1 domain";  
 RL J. Immunol. 147:174-180(1991).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M62852; AAA59647.1; -.  
 DR HSP: P03989; IHSA.  
 DR MIN: 142830; -.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC.I; 1.  
 KW MHC I; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 90 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 91 182 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 183 274 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 275 284 CONNECTING PEPTIDE.  
 FT TRANSMEM 285 308  
 FT DOMAIN 309 338 CYTOPLASMIC TAIL.  
 FT CARBOHYD 86 86 BY SIMILARITY.  
 FT DISULFID 101 164 BY SIMILARITY.  
 FT DISULFID 203 259 BY SIMILARITY.  
 SQ SEQUENCE 338 AA; 37804 MW; 33FB8134 CRC32;

Query Match 88.7%; Score 118; DB 1; Length 338;  
 Best Local Similarity 88.0%; Pred. No. 9e-11;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 DB 60 WDRETOICKAKAQTDRNLRIALRY 84







```

RESULT 7
1B19_HUMAN
ID ID 1B19_HUMAN STANDARD; PRT; 362 AA.
AC Q08136;
DT DT 01-NOV-1995 (Rel. 32, Created)
DT DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2706 ALPHA CHAIN
DE DE PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.A.
RX MEDLINE: 94102824.
RA VILCHES C., DE PABLO R., KREISLER M.;
RT "Nucleotide sequence of HLA-B*2706." ;
RL Immunogenetics 39:219-219(1994).
RN [2]
RP SEQUENCE FROM N.A.A.
RX MEDLINE: 96134006.
RA RUDELWALT M., BOWNESS P., WORDSWORTH P.;
RT "The nucleotide sequence of HLA-B*2704 reveals a new amino acid
RT substitution in exon 4 which is also present in HLA-B*2706." ;
RL Immunogenetics 43:160-162(1996).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to licensedsib-sib.ch).
CC -----
EMBL; X73578; CAAS1980.1; -
DR EMBL; U35734; AAC50447.1; -.
DR HSSP; P03989; IHSA.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL
FT CHAIN 1 24
FT CHAIN 25 362
HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
B-27 B*2706 ALPHA CHAIN.
EXTRACELLULAR ALPHA-1 (BY SIMILARITY).
EXTRACELLULAR ALPHA-2 (BY SIMILARITY).
EXTRACELLULAR ALPHA-3 (BY SIMILARITY).
CONNECTING PEPTIDE (BY SIMILARITY).
BY SIMILARITY.
CYTOPLASMIC TAIL (BY SIMILARITY).
POTENTIAL.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
A -> G (IN REF. 1).
SEQUENCE 362 AA; 04056 MW; 0D402027 CRC32;

Query Match 88.7%; Score 118; DB 1; Length 362;
Best Local Similarity 88.0%; Pred. NO. 9.6e-11;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETIICKAKAQTDRENRIALRY 25
|||||
DB 84 WDRETIICKAKAQTDRESLIIRLY 108
|||||

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE: 92078860.
RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT "Gorilla class I major histocompatibility complex alleles: comparison
to human and chimpanzee class I.";
RL J. Exp. Med. 174:1491-1509(1991).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60593; CAA43101.1; -.
DR PIR: JH0540; JH0540.
DR HSSP: P03989; IHSA.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT CLASS I HISTOCOMPATIBILITY ANTIGEN,
GOGO-B0103 ALPHA CHAIN.
FT DOMAIN 25 114
EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206
EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298
EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308
CONNECTING PEPTIDE.
FT TRANSMEM 309 332
CYTOPLASMIC TAIL.
FT DOMAIN 333 362
BY SIMILARITY.
FT DISULFID 125 188
BY SIMILARITY.
FT DISULFID 227 283
BY SIMILARITY.
FT CARBOHYD 110 110
BY SIMILARITY.
SQ SEQUENCE 362 AA; 40204 MW; 3CF119AD CRC32;

Query Match 85.7%; Score 114; DB 1; Length 362;
Best Local Similarity 88.0%; Pred. No. 3.9e-10;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25
||||| ||:|||||
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 10
ID IB03_GORGO STANDARD; PRT; 362 AA.
AC P30381;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE "Gorilla class I major histocompatibility complex alleles: comparison
to human and chimpanzee class I.";
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92078860.
RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT "Gorilla class I major histocompatibility complex alleles: comparison
to human and chimpanzee class I.";
RL J. Exp. Med. 174:1491-1509(1991).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60593; CAA43101.1; -.
DR PIR: JH0540; JH0540.
DR HSSP: P03989; IHSA.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT CLASS I HISTOCOMPATIBILITY ANTIGEN,
GOGO-B0103 ALPHA CHAIN.
FT DOMAIN 25 114
EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206
EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298
EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308
CONNECTING PEPTIDE.
FT TRANSMEM 309 332
CYTOPLASMIC TAIL.
FT DOMAIN 333 362
BY SIMILARITY.
FT DISULFID 125 188
BY SIMILARITY.
FT DISULFID 227 283
BY SIMILARITY.
FT CARBOHYD 110 110
BY SIMILARITY.
SQ SEQUENCE 362 AA; 40204 MW; 3CF119AD CRC32;

Query Match 85.7%; Score 114; DB 1; Length 362;
Best Local Similarity 88.0%; Pred. No. 3.9e-10;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25
||||| ||:|||||
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 10
ID IB03_GORGO STANDARD; PRT; 362 AA.
AC P30381;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE "Gorilla class I major histocompatibility complex alleles: comparison
to human and chimpanzee class I.";
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92078860.
RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT "Gorilla class I major histocompatibility complex alleles: comparison
to human and chimpanzee class I.";
RL J. Exp. Med. 174:1491-1509(1991).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60593; CAA43101.1; -.
DR PIR: JH0540; JH0540.
DR HSSP: P03989; IHSA.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT CLASS I HISTOCOMPATIBILITY ANTIGEN,
GOGO-B0103 ALPHA CHAIN.
FT DOMAIN 25 114
EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206
EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298
EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308
CONNECTING PEPTIDE.
FT TRANSMEM 309 332
CYTOPLASMIC TAIL.
FT DOMAIN 333 362
BY SIMILARITY.
FT DISULFID 125 188
BY SIMILARITY.
FT DISULFID 227 283
BY SIMILARITY.
FT CARBOHYD 110 110
BY SIMILARITY.
SQ SEQUENCE 362 AA; 40204 MW; 3CF119AD CRC32;

Query Match 85.7%; Score 114; DB 1; Length 362;
Best Local Similarity 88.0%; Pred. No. 3.9e-10;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25
||||| ||:|||||
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 11
ID IB01_PANTR STANDARD; PRT; 359 AA.
AC P13750;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN PRECURSOR
(FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89030641.
RA MAYER W.E., JONKER M., KLEIN D., IVANYI P., VAN SEVENTER G.,
KLEIN J.;
RT "Nucleotide sequences of chimpanzee MHC class I alleles: evidence for
trans-species mode of evolution.";
RL EMBO J. 7:2765-2774(1988).
RN [2]
RP REVISIONS.
RA MAYER W.;
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60254; CAA42806.1; -.
DR PIR: JH0541; JH0541.
DR HSSP: P03989; IHSA.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT CLASS I HISTOCOMPATIBILITY ANTIGEN,
GOGO-B0103 ALPHA CHAIN.
FT DOMAIN 25 114
EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206
EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298
EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308
CONNECTING PEPTIDE.
FT TRANSMEM 309 332
CYTOPLASMIC TAIL.
FT DOMAIN 333 362
BY SIMILARITY.
FT DISULFID 125 188
BY SIMILARITY.
FT DISULFID 227 283
BY SIMILARITY.
FT CARBOHYD 110 110
BY SIMILARITY.
SQ SEQUENCE 362 AA; 40248 MW; FEA6A941 CRC32;

Query Match 85.7%; Score 114; DB 1; Length 362;
Best Local Similarity 88.0%; Pred. No. 3.9e-10;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25
||||| ||:|||||
Db 84 WDRETOICKAKAQTDRNLRIALRY 108

RESULT 11
ID IB01_PANTR STANDARD; PRT; 359 AA.
AC P13750;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN PRECURSOR
(FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89030641.
RA MAYER W.E., JONKER M., KLEIN D., IVANYI P., VAN SEVENTER G.,
KLEIN J.;
RT "Nucleotide sequences of chimpanzee MHC class I alleles: evidence for
trans-species mode of evolution.";
RL EMBO J. 7:2765-2774(1988).
RN [2]
RP REVISIONS.
RA MAYER W.;
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; X13115; CAA31507.1; -.  
 CC DR PIR; S03537; S03537.  
 CC DR HSP; P03989; IHS.  
 CC DR PROSITE; PS00290; IG\_MHC; 1.  
 CC DR PFAM; PF00047; Iq; 1.  
 CC DR PFAM; PF00129; MHC\_I; 1.  
 CC KW MHC I; transmembrane; Glycoprotein; Signal.  
 CC FT NON\_TER 1 1  
 CC FT SIGNAL <1 20  
 CC FT CHAIN 21 359  
 CC FT DOMAIN 21 110  
 CC FT DOMAIN 111 202  
 CC FT DOMAIN 203 294  
 CC FT DOMAIN 295 305  
 CC FT TRANSMEM 306 329  
 CC FT DOMAIN 330 359  
 CC FT DISULFID 121 184  
 CC FT DISULFID 223 279  
 CC FT CARBOHYD 106 106  
 CC FT SEQUENCE 359 AA; 5395FFC9 CRC32;  
 CC SQ

Query Match 80.58; Score 107; DB 1; Length 359;  
 Best Local Similarity 84.08; Pred. No. 4.5e-09;  
 Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLALRY 25  
 ||||| || ||||| ||||| |||||  
 DB 80 WDRETRNKASQTDRENRLALRY 104

RESULT 12  
 1B47\_HUMAN  
 ID 1B47\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30487;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-49(B-21) B\*4901 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89235215.  
 RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
 RT "Diversity and diversification of HLA-A,B,C alleles."  
 RL J. Immunol. 142:3937-3950(1989).  
 RN [2]  
 RP REVISION TO 78.  
 RX MEDLINE; 93056529.  
 RA HILDEBRAND W.H., MADRIGAL J.A., BELICH M.P., ZEMMOUR J., WARD F.E.,  
 RA WILLIAMS R.C., PARHAM P.;  
 RT "Serologic cross-reactivities poorly reflect allelic relationships in  
 RT the HLA-B12 and HLA-B21 groups. Dominant epitopes of the alpha 2  
 RT helix."  
 RL J. Immunol. 149:3563-3568(1992).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; M24037; AAA02950.1; -.  
 CC DR HSP; P30491; 1A1M.  
 CC DR MIM; 142830; -.  
 CC DR PROSITE; PS00290; IG\_MHC; 1.  
 CC DR PFAM; PF00047; Iq; 1.  
 CC DR PFAM; PF00129; MHC\_I; 1.  
 CC KW MHC I; transmembrane; Glycoprotein; Signal.  
 CC FT SIGNAL 1 24  
 CC FT CHAIN 25 362  
 CC FT DOMAIN 25 114  
 CC FT DOMAIN 115 206  
 CC FT DOMAIN 207 298  
 CC FT DOMAIN 299 309  
 CC FT TRANSMEM 310 333  
 CC FT DOMAIN 334 362  
 CC FT CARBOHYD 110 110  
 CC FT DISULFID 125 188  
 CC FT DISULFID 227 283  
 CC FT SEQUENCE 362 AA; 40581 MW; E996F82F CRC32;  
 CC SQ

Query Match 75.9%; Score 101; DB 1; Length 362;  
 Best Local Similarity 80.08; Pred. No. 3.7e-08;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLALRY 25  
 ||||| || ||||| ||||| |||||  
 DB 84 WDRETOISKNTQTYRENRLALRY 108

RESULT 13  
 1B53\_HUMAN  
 ID 1B53\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30490;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-52(B-5) B\*5201 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89080265.  
 RA HAYASHI H., ENNIS P.D., ARIGA H., SALTER R.D., PARHAM P., KANO K.,  
 RA TAKIGUCHI M.;  
 RT "HLA-B51 and HLA-Bw52 differ by only two amino acids which are in the  
 RT helical region of the alpha 1 domain."  
 RL J. Immunol. 142:306-311(1989).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M22799; AAA59645.1; ALT\_SEQ.  
 CC EMBL; M22793; AAA59645.1; JOINED.  
 CC EMBL; M22794; AAA59645.1; JOINED.  
 CC EMBL; M22795; AAA59645.1; JOINED.  
 CC EMBL; M22796; AAA59645.1; JOINED.

```

DR EMBL: M22797; AAA59645.1; JOINED.
DR EMBL: M22798; AAA59645.1; JOINED.
DR PIR: B30345; B30345.
DR PIR: B30548; B30548.
DR HSP: P30491; IAIM.
DR MIM: 142830; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 308
FT TRANSMEM 309 332
FT DOMAIN 333 362
FT CARBOHYD 110 110
FT DISULFID 125 188
FT DISULFID 227 283
FT SEQUENCE 362 AA; 40521 MW; 38436FE8 CRC32;

Query Match 75.9%; Score 101; DB 1; Length 362;
Best Local Similarity 80.0%; Pred. No. 3.7e-08;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WDRETIQCKAKAQTDRENLRIALRY 25
Db 84 WDRETIQSKTNTQTYRENLRIALRY 108

RESULT 14
1B05_HUMAN
ID 1B05_HUMAN STANDARD; PRT; 362 AA.
AC P30461;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-13 B*1301 ALPHA CHAIN
DE PRECURSOR (B13.1).
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89235215.
RA PARHAM P., LAWLOR D.A., LOMEN C.E., ENNIS P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE: 88152906.
RA ZEMMOUR J., ENNIS P.D., PARHAM P., DUPONT B.;
RT "Comparison of the structure of HLA-Bw47 to HLA-B13 and its
RT relationship to 21-hydroxylase deficiency.";
RL Immunogenetics 27:281-287(1988).
[3]
RN SEQUENCE FROM N.A.
RX TISSUE-BLOOD.
RX MEDLINE: 96053518.
RA LIN L., TOKUNAGA K., NAKAJIMA F., ISHIKAWA Y., KASHIWASE K.,
RA TANAKA H., KUWATA S., SIDAITSOVA E., AKAZA T., TABOKORO K.,
RA SHIBATA Y., CHANDANAYINGYONG D., JUJI T.;
RT "Both HLA-B*1301 and B*1302 exist in Asian populations and are
RT associated with different haplotypes.";
RL Hum. Immunol. 43:51-56(1995).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
-----
MICROGLOBULIN).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: M24041; AAA59660.1; -.
DR EMBL: M19757; AAA52657.1; -.
DR EMBL: D50291; BRA08822.1; -.
DR HSP: P30491; IAIM.
DR MIM: 142830; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 309
FT TRANSMEM 310 333
FT DOMAIN 334 362
FT CARBOHYD 110 110
FT DISULFID 125 188
FT DISULFID 227 283
FT SEQUENCE 362 AA; 40474 MW; 28867875 CRC32;

Query Match 72.2%; Score 96; DB 1; Length 362;
Best Local Similarity 76.0%; Pred. No. 2.1e-07;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WDRETIQCKAKAQTDRENLRIALRY 25
Db 84 WDRETIQSKTNTQTYRENLRIALRY 108

RESULT 15
1B41_HUMAN
ID 1B41_HUMAN STANDARD; PRT; 362 AA.
AC P30481;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-44(B-12) B*4402 ALPHA CHAIN
DE PRECURSOR (B44.2).
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89235215.
RA PARHAM P., LAWLOR D.A., LOMEN C.E., ENNIS P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
-----
MICROGLOBULIN).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@sib-sib.ch).

CC -----  
 DR EMBL: M24038; AAA59663.1; -  
 DR HSP; P30491; IAIM.  
 DR MIM; 142830; -  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 BW-44(B-12) B\*4402 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 309 CONNECTING PEPTIDE.  
 FT TRANSMEM 310 333 CYTOPLASMIC TAIL.  
 FT DOMAIN 334 362 BY SIMILARITY.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40481 MW; D0AE6DD5 CRC32;

Query Match 72.2%; Score 96; DB 1; Length 362;

Best Local Similarity 76.0%; Pred. No. 2.1e-07;

Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 WDRETOICKAKAQTDRNLRLRY 25  
 ||||| | |||||  
 Db 84 WDRETOISKNTQTYRNLRLRY 108

Search completed: February 8, 2000, 01:26:01  
 Job time: 1561 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 19:16:17 ; Search time 176.54 seconds  
(without alignments)  
9.818 Million cell updates/sec

Title: US-08-653-294-32

Perfect score: 133

Sequence: 1 WDRETQICKAKAQTDRNLRLRY 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPSTREMBL\_12.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_unclassified.\*  
14: sp\_unclassified.\*

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	133	100.0		133	7	019189	019189 homo sapien
2	124	93.2		354	7	Q95524	Q95524 pan paniscu
3	124	93.2		354	7	Q95527	Q95527 pan paniscu
4	120	90.2		90	7	O46697	O46697 gorilla gor
5	118	88.7		39	7	O19688	019688 homo sapien
6	118	88.7		90	7	O19193	019193 homo sapien
7	118	88.7		181	7	O78142	078142 homo sapien
8	118	88.7		274	7	O19692	019692 homo sapien
9	118	88.7		359	7	Q29934	Q29934 homo sapien
10	118	88.7		362	7	Q29693	Q29693 homo sapien
11	118	88.7		362	7	Q29846	Q29846 homo sapien
12	118	88.7		362	7	P79612	P79612 homo sapien
13	118	88.7		362	7	O78189	O78189 homo sapien
14	117	88.0		90	7	O46693	O46693 pan troglod
15	117	88.0		90	7	O46694	O46694 pan troglod
16	117	88.0		90	7	O46695	O46695 pan troglod
17	117	88.0		90	7	O46696	O46696 pan paniscu
18	114	85.7		181	7	O19354	O19354 gorilla gor
19	107	80.5		354	7	Q95412	Q95412 hylobates l
20	107	80.5		355	7	Q95529	Q95529 pan troglod

## ALIGNMENTS

```

RESULT      1
ID          019189
ID ID       PRELIMINARY; PRT; 133 AA.
AC AC
DT DT 01-JAN-1998 (TReMBUrel. 05, Created)
DT DT 01-JAN-1998 (TReMBUrel. 05, Last sequence update)
DT DT 01-NOV-1999 (TReMBUrel. 12, Last annotation update)
DE DE MHC CLASS I HISTOCOMPATIBILITY ANTIGEN-B (HLA-B-27KSH) (FRAGMENT).
GN HLA-B.
ON ON Homo sapiens (Human).
GC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
[1] [1] SEQUENCE FROM N.A.
RN RN TISSUE=LEUKOCYTE;
RC RC PETERSDORF E.;
RA RA Submitted (DEC-1994) to the EMBL/GenBank/DDJB databases.
DR DR EMBL; U18659; AAB60357.1; -.
DR DR MIM; 142830; -.
DR DR PFAM; PF00129; MHC_I; 1.
KW KW MHC I.
FT FT NON_TER 1 1
FT FT NON_TER 133 133
SQ SQ SEQUENCE 133 AA; 15491 MW; 3A3BC802 CRC32;

Query Match 100.0%; Score 133; DB 7; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.le-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY QY 1 WDRETQICKAKAQTDRENRLALRY 25
    |||||
DB DB 12 WDRETOICKAKAQTDRENRLALRY 36
    |||||

RESULT      2
ID          Q95524
ID ID       PRELIMINARY; PRT; 354 AA.
AC AC
DT DT 01-FEB-1997 (TReMBUrel. 02, Created)
DT DT 01-FEB-1997 (TReMBUrel. 02, Last sequence update)
DT DT 01-NOV-1999 (TReMBUrel. 12, Last annotation update)
DE DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).

```

GN HLA-B.  
 OS Pan paniscus (Pygmy chimpanzee) (Bonobo).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LOREL;  
 RX MEDLINE; 94286544.  
 RA MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,  
 RA BONTROP R.E., WATKINS D.I.;  
 RT "A uniquely high level of recombination at the HLA-B locus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5893-5897(1994).  
 DR EMBL; U05575; AAA50178.1; -;  
 DR HSSP; P30491; IALM.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 354 AA; 39227 MW; 5D75939D CRC32;

Query Match 93.2%; Score 124; DB 7; Length 354;  
 Best Local Similarity 92.0%; Pred. No. 1.4e-11;  
 Matches 23; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRLRY 25  
 ||| |||||:|||||||  
 Db 76 WDRNTQICKAKAQTDRNLRLRY 100

RESULT 3  
 Q95527 PRELIMINARY; PRT; 354 AA.  
 AC Q95527;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).  
 GN HLA-B.  
 OS Pan paniscus (Pygmy chimpanzee) (Bonobo).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BOSONDJO;  
 RX MEDLINE; 94286544.  
 RA MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,  
 RA BONTROP R.E., WATKINS D.I.;  
 RT "A uniquely high level of recombination at the HLA-B locus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5893-5897(1994).  
 DR EMBL; U05578; AAA50181.1; -;  
 DR HSSP; P30491; IALM.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 354 AA; 39401 MW; 6E38CFFE CRC32;

Query Match 93.2%; Score 124; DB 7; Length 354;  
 Best Local Similarity 92.0%; Pred. No. 1.4e-11;  
 Matches 23; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRLRY 25  
 ||| |||||:|||||||  
 Db 76 WDRNTQICKAKAQTDRNLRLRY 100

RESULT 4  
 O46697 PRELIMINARY; PRT; 90 AA.  
 ID O46697

AC O46697;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-H ORTHOLOG (FRAGMENT).  
 GN HLA-H.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SHAMBA;  
 RA GRIMSLEY C., MATHER K.A., OBER C.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022172; AAC99794.1; -;  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 90 90  
 SQ SEQUENCE 90 AA; 10689 MW; 5E5F2495 CRC32;

Query Match 90.2%; Score 120; DB 7; Length 90;  
 Best Local Similarity 88.0%; Pred. No. 1.5e-11;  
 Matches 22; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRLRY 25  
 ||| |||||:|||||||  
 Db 60 WDRNTQICKAKAQTDRNLRLRY 84

RESULT 5  
 O19688 PRELIMINARY; PRT; 39 AA.  
 AC O19688;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE HLA-B\*27 VARIANT EXON 2 (ALPHA 1 DOMAIN) (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BLASZYK R., WEBER M., SALAVA A.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X83727; CAA58698.1; -;  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 39 39  
 SQ SEQUENCE 39 AA; 4748 MW; 6F714D4C CRC32;

Query Match 88.7%; Score 118; DB 7; Length 39;  
 Best Local Similarity 88.0%; Pred. No. 1.3e-11;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WDRETOICKAKAQTDRNLRLRY 25  
 ||| |||||:|||||||  
 Db 9 WDRNTQICKAKAQTDRNLRLRY 33

RESULT 6  
 O19193 PRELIMINARY; PRT; 90 AA.  
 AC O19193;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE HISTOCOMPATIBILITY ANTIGEN ALPHA 1 DOMAIN (FRAGMENT).  
 GN HLA-B\*27.  
 OS Homo sapiens (Human).



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 9237445.  
 RA HIGGINS C.M., LUND T., SHIPLEY M.E., EBRINGER A.,  
 RA SADOWSKA-WROBLEWSKA M., CRAIG R.K.;  
 RT "Ankylosing spondylitis and HLA-B\*27: restriction fragment length  
 RT polymorphism and sequencing of an HLA-B\*27 allele from a patient with  
 RT ankylosing spondylitis.";  
 RL Ann. Rheum. Dis. 51:855-862(1992).  
 DR EMBL: S39758; CAB27364.1; -;  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 90 90  
 SQ SEQUENCE 90 AA; 10571 MW; F22CCB4E CRC32;

Query Match 88.7%; Score 118; DB 7; Length 90;  
 Best Local Similarity 88.0%; Pred. No. 3e-11;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 60 WDRETOICKAKAQTDRDLRLTLLRY 84

RESULT 7  
 O78142  
 ID O78142 PRELIMINARY; PRT; 181 AA.  
 AC O78142;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE MHC CLASS I ANTIGEN (FRAGMENT).  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STEINER N.K., HURLEY C.K., KOESTER R.P.;  
 RT "Novel-HLA-B allele";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF072764; AAC25779.1; -;  
 DR EMBL: AF072763; AAC25779.1; JOINED.  
 DR HSSP: P10318; IROG.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 181 181  
 SQ SEQUENCE 181 AA; 21079 MW; 24949B0F CRC32;

Query Match 88.7%; Score 118; DB 7; Length 181;  
 Best Local Similarity 88.0%; Pred. No. 6.1e-11;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 59 WDRETOICKAKAQTDRDLRLTLLRY 83

RESULT 8  
 O19692  
 ID O19692 PRELIMINARY; PRT; 274 AA.  
 AC O19692;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE MHC CLASS I HLA-B\*27 M (FRAGMENT).  
 GN HLA-B.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87009855.  
 RA COPPIN H.L., MCDEVITT H.O.;  
 RT "Absence of polymorphism between HLA-B\*27 genomic exon sequences  
 RT isolated from normal donors and ankylosing spondylitis patients.";  
 RL J. Immunol. 137:2168-2172(1986).  
 DR EMBL: M14013; AAA59643.1; -;  
 DR HSSP: P10318; IROG.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 274 274  
 SQ SEQUENCE 274 AA; 31659 MW; 9A74A6BA CRC32;

Query Match 88.7%; Score 118; DB 7; Length 274;  
 Best Local Similarity 88.0%; Pred. No. 9.2e-11;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 60 WDRETOICKAKAQTDRDLRLTLLRY 84

RESULT 9  
 Q29934  
 ID Q29934 PRELIMINARY; PRT; 359 AA.  
 AC Q29934;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE HLA-B\*27 (FRAGMENT).  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86149317.  
 RA SZOIS H., RIEHMULLER G., WEISS E., MEO T.;  
 RT "Complete sequence of HLA-B\*27 cDNA identified through the  
 RT characterization of structural markers unique to the HLA-A, -B, and -C  
 RT allelic series.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1428-1432(1986).  
 DR EMBL: M12678; AAA59614.1; -;  
 DR HSSP: P10318; IROG.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 359 AA; 40042 MW; 069F7E64 CRC32;

Query Match 88.7%; Score 118; DB 7; Length 359;  
 Best Local Similarity 88.0%; Pred. No. 1.2e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 |||||  
 Db 81 WDRETOICKAKAQTDRDLRLTLLRY 105

RESULT 10  
 Q29693  
 ID Q29693 PRELIMINARY; PRT; 362 AA.  
 AC Q29693;

DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE MHC CLASS I ANTIGEN.  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 96134006.  
 RA RUDWALEIT M., BOWNESS P., WORDSWORTH P.;  
 RT "The nucleotide sequence of HLA-B\*2704 reveals a new amino acid  
 RT substitution in exon 4 which is also present in HLA-B\*2706."  
 RL Immunogenetics 43:160-162(1996).  
 DR EMBL; U27608; AAC50444.1; -.  
 DR HSSP; P10318; IROG.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 SQ SEQUENCE 362 AA; 40416 MW; 850432B4 CRC32;

Query Match 88.7%; Score 118; DB 7; Length 362;  
 Best Local Similarity 88.0%; Pred. No. 1.2e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLRY 25  
 |||||  
 DB 84 WDRETOICKAKAQTDRNLRLRY 108

## RESULT 11

ID Q29846 PRELIMINARY; PRT; 362 AA.  
 AC Q29846;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE HUMAN LYMPHOCYTE ANTIGEN HLA-B\*27.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-BLOOD;  
 RX MEDLINE; 94375872.  
 RA DEL PORTO P., D'AMATO M., FIORILLO M.T., TUOSTO L., PICCOLELLA E.,  
 RA SORRENTINO R.;  
 RT "Identification of a novel HLA-B\*27 subtype by restriction analysis of  
 RT a cytotoxic gamma delta T cell clone."  
 RL J. Immunol. 153:3093-3100(1994).  
 DR EMBL; Z33453; CAA83876.1; -.  
 DR HSSP; P10318; IROG.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 SQ SEQUENCE 362 AA; 40450 MW; CCA23A50 CRC32;

Query Match 88.7%; Score 118; DB 7; Length 362;  
 Best Local Similarity 88.0%; Pred. No. 1.2e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLRY 25  
 |||||  
 DB 84 WDRETOICKAKAQTDRNLRLRY 108

## RESULT 12

P79612

ID P79612 PRELIMINARY; PRT; 362 AA.  
 AC P79612;  
 DT 01-MAY-1997 (TREMELrel. 03, Created)  
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE HLA-B PROTEIN.  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-BLOOD;  
 RA HASEGAWA T., SUGAHARA Y., MORIYAMA Y., NANZAI H., OGAWA A., TAWARA K.,  
 RA KONDO S., TOKUNAGA K.;  
 RT "Molecular characterization of a novel HLA-B\*27 allele."  
 RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D83043; BAAL1753.1; -.  
 DR HSSP; P10318; IROG.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 SQ SEQUENCE 362 AA; 40441 MW; 1D8099BC CRC32;

Query Match 88.7%; Score 118; DB 7; Length 362;  
 Best Local Similarity 88.0%; Pred. No. 1.2e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLRY 25  
 |||||  
 DB 84 WDRETOICKAKAQTDRNLRLRY 108

## RESULT 13

ID 078189 PRELIMINARY; PRT; 362 AA.  
 AC 078189;  
 DT 01-NOV-1998 (TREMELrel. 08, Created)  
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE MHC CLASS I ANTIGEN.  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RA SEURYNCK K.L., BAXTER-LOWE L.A.;  
 RT "B27052 W496D."  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF026218; AAC42275.1; -.  
 DR HSSP; P10318; IROG.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 SQ SEQUENCE 362 AA; 40486 MW; 2B0EF602 CRC32;

Query Match 88.7%; Score 118; DB 7; Length 362;  
 Best Local Similarity 88.0%; Pred. No. 1.2e-10;  
 Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRLRY 25  
 |||||  
 DB 84 WDRETOICKAKAQTDRNLRLRY 108

## RESULT 14

O46693 PRELIMINARY; PRT; 90 AA.  
 ID O46693  
 AC O46693;

DT 01-JUN-1998 (TReMBLrel. 06, Created)  
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-H ORTHOLOG (FRAGMENT).  
 GN HLA-H.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ROSS;  
 RA GRIMSLEY C., MATHER K.A., OBER C.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022168; AAC99790.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 90 90  
 SQ SEQUENCE 90 AA; 10707 MW; EE865717 CRC32;

Query Match 88.0%; Score 117; DB 7; Length 90;  
 Best Local Similarity 84.0%; Pred. No. 4.3e-11;  
 Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 ||| |||||:||||:||||:||||  
 Db 60 WDRNTQICKAQAQTERNLRIALRY 84

RESULT 15  
 O46694  
 ID O46694 PRELIMINARY; PRT: 90 AA.  
 AC O46694;  
 DT 01-JUN-1998 (TReMBLrel. 06, Created)  
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-H ORTHOLOG (FRAGMENT).  
 GN HLA-H.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EVE;  
 RA GRIMSLEY C., MATHER K.A., OBER C.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022169; AAC99791.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 90 90  
 SQ SEQUENCE 90 AA; 10707 MW; EE865717 CRC32;

Query Match 88.0%; Score 117; DB 7; Length 90;  
 Best Local Similarity 84.0%; Pred. No. 4.3e-11;  
 Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 WDRETOICKAKAQTDRNLRIALRY 25  
 ||| |||||:||||:||||:||||  
 Db 60 WDRNTQICKAQAQTERNLRIALRY 84

Search completed: February 8, 2000, 19:16:17  
 Job time: 21504 sec

**THIS PAGE BLANK (USPTO)**

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation	..
gb.pr2:HSU18559	+	133.00	402.28	2.2e-14	399	L18659 Human MHC class I HLA-B*	
gb.pr2:HUHLABC	+	133.00	393.67	6.6e-14	1089	L38504 Homo sapiens (clones 18	
gb.pr1:HSHLAK1	+	124.00	392.55	7.7e-14	1241	X03664 Human Class I MHC gene	
gb.pr1:PPU05575	+	123.00	365.73	2.4e-12	1065	U05575 Pan paniscus Class I hla	
gb.pr1:PPU05578	+	124.00	365.73	2.4e-12	1065	U05578 Pan paniscus Class I hla	
gb.pr1:HSHLABHA	+	120.00	365.65	2.4e-12	250	X08692 H.sapiens HLA-B gene, ex	
gb.pr1:HSHLABHB	+	120.00	365.65	2.4e-12	250	X08693 H.sapiens HLA-B gene, ex	
gb.pr4:AF022159	+	120.00	365.48	2.5e-12	255	AF022159 Homo sapiens isolate Q	
gb.pr4:AF022161	+	120.00	364.96	2.6e-12	271	AF022161 Homo sapiens isolate Q	
gb.pr4:AF022163	+	120.00	364.96	2.6e-12	271	AF022163 Homo sapiens isolate Q	
gb.pr4:AF022164	+	120.00	364.96	2.6e-12	271	AF022164 Homo sapiens isolate Q	
gb.pr4:AF022165	+	120.00	364.96	2.6e-12	271	AF022165 Homo sapiens isolate Q	
gb.pr4:AF022166	+	120.00	364.96	2.6e-12	271	AF022166 Homo sapiens isolate Q	
gb.pr4:AF022167	+	120.00	364.96	2.6e-12	271	AF022167 Gorilla gorilla isolate	
gb.pr2:HUHHACR3	+	120.00	352.98	1.2e-11	1096	M32106 Human MHC HLA-AR gene,	
gb.pr1:HUHHAC1R	+	120.00	341.62	5.3e-11	4123	J00191 Human MHC Class I trans	
gb.pr1:HSMHNT1	+	120.00	340.80	5.8e-11	4537	X12432 Human major histocompat	
gb.pr1:HUHHANTLE	+	120.00	337.62	8.8e-11	6581	M31944 Human MHC class I proc	
gb.pr3:AF050566	-	120.00	304.33	6.3e-09	319486	AF050566 Homo sapiens MHC cl	
gb.pr2:HUMB270	+	119.00	355.83	8.5e-12	546	L76935 Homo sapiens MHC class I	
gb.pr2:HSUAB27V2	+	118.00	361.53	4.4e-12	195	X83727 H.sapiens HLA-B*7 varian	
gb.pr3:HSAB27HLA1	+	118.00	358.74	5.8e-12	270	AF027263 Homo sapiens MHC class	
gb.pr4:HSHLAWNB2	+	118.00	358.74	5.8e-12	270	AF110257 Homo sapiens MHC class	
gb.pr4:HSMLCCCHN1	+	118.00	358.74	5.8e-12	270	AF170578 Homo sapiens MHC class	
gb.pr2:HSMLCCCHN2	+	118.00	358.74	5.8e-12	270	AF170578 Homo sapiens MHC class	
gb.pr2:HUMB2705G	+	118.00	352.70	1.3e-11	546	L76095 Homo sapiens MHC class	
gb.pr1:HUHHB2	+	118.00	349.20	2.0e-11	822	M14013 Human MHC class I HLA-B*	
gb.pr1:HUHHB2C	+	118.00	347.37	2.5e-11	1017	M62852 Human MHC class I HLA-B*	
gb.pat:AS28264	+	118.00	347.30	2.5e-11	1026	A28264 H.sapiens mRNA for HLA-	
gb.pat:E01342	+	118.00	347.30	2.5e-11	1026	E01342 cDNA encoding C-termina	
gb.pr1:HUHHB2UN1A	+	118.00	346.82	2.7e-11	1084	L20086 Human MHC class I (HLA-	
gb.pr1:HSHLABA	+	118.00	346.79	2.7e-11	1089	X73578 H.sapiens HLA-B*2705 m	
gb.pr3:HS0276218	+	118.00	346.79	2.7e-11	1089	AF026218 Homo sapiens MHC class	
gb.pr3:HS027608	+	118.00	346.79	2.7e-11	1089	AF02608 Homo sapiens class I anti	
gb.pr3:HSU35734	+	118.00	346.79	2.7e-11	1089	X95734 Human HLA-B class I anti	
gb.pr1:DS83043	+	118.00	346.70	2.7e-11	1100	DS8043 Human HLA-B mRNA, allel	
gb.pr1:HSHLAW1	+	118.00	345.66	3.1e-11	1241	X03665 Human Class I MHC gene	
gb.pr1:HSHLAB27X	+	118.00	345.45	3.2e-11	1272	Z33453 H.sapiens mRNA for huma	
gb.pr1:HUHHB27A	+	118.00	344.00	3.9e-11	1507	M12678 Human HLA-B*27 mRNA, com	
gb.pr1:HUHHB27D	+	118.00	336.42	1.0e-10	3649	M34883 Human MHC class I HLA-B*	
gb.pat:E01341	+	118.00	335.60	1.1e-10	4015	E01341 Genomic DNA encoding HLA	
gb.pr1:HUHHB27B	+	118.00	335.60	1.1e-10	4016	M12967 Human MHC class I HLA-B*	
gb.pr4:IS39758	+	118.00	335.13	1.2e-10	4242	X39758 HLA-B*27 HLA-B*2705s-h	









```

pseudogene, partial sequence.
ACCESSION AF022159
VERSION AF022159.1 GI:2655062
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 255)
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE HLA-H: a pseudogene with increased variation due to balancing
JOURNAL Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE 99083426
FEATURES
    Location/Qualifiers
        1..255
            /organism="Homo sapiens"
            /isolate="026"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6p21.3"
            /note="African-American individual"
        exon
            1..255
            /gene="HLA-H"
            /pseudo
            /number=2
            <1..>255
            /note="HLA-H"
            /note="MHC Class I antigen HLA-H"
            /codon_start=1
            /pseudo
            <1..>255
            /gene="HLA-H"
            /pseudo
        CDS
            50 a 81 c 86 g 38 t
            alignment_scores:
                Quality: 120.00 Length: 25
                Ratio: 4.800 Gaps: 0
            Percent Similarity: 100.000 Percent Identity: 88.000
        alignment_block:
            US-08-653-294-32 x AF022159 ..
            Align seg 1/1 to: AF022159 from: 1 to: 255
            1 TrpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
              |||||:::|||||:::|||||:::|||||:::|||||
            178 TGGACCGGAACACACAGATCTGCAAGCCGACGACGACTGAACGAGA 227

17 uAsnLeuArgIleAlaLeuArgTyr 25
228 GAACCTCGGATCGGCTCGCTAC 252

seq_name: gb_pr4:AF022161

seq_documentation_block:
LOCUS AF022161 271 bp DNA PRI 05-JAN-1999
DEFINITION Homo sapiens isolate 068 MHC class I antigen HLA-H (HLA-H)
pseudogene, partial sequence.
ACCESSION AF022161
VERSION AF022161.1 GI:2655064
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE HLA-H: a pseudogene with increased variation due to balancing
JOURNAL Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE 99083426

```

```

Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE HLA-H: a pseudogene with increased variation due to balancing
JOURNAL Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE 99083426
FEATURES
    Location/Qualifiers
        1..271
            /organism="Homo sapiens"
            /isolate="068"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6p21.3"
            /note="African-American individual"
        exon
            1..270
            /gene="HLA-H"
            /pseudo
            /number=2
            <1..>270
            /note="HLA-H"
            /note="MHC Class I antigen HLA-H"
            /codon_start=1
            /pseudo
            <1..>271
            /gene="HLA-H"
            /pseudo
        CDS
            56 a 86 c 92 g 37 t
            alignment_scores:
                Quality: 120.00 Length: 25
                Ratio: 4.800 Gaps: 0
            Percent Similarity: 100.000 Percent Identity: 88.000
        alignment_block:
            US-08-653-294-32 x AF022161 ..
            Align seg 1/1 to: AF022161 from: 1 to: 271
            1 TrpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
              |||||:::|||||:::|||||:::|||||:::|||||
            178 TGGACCGGAACACACAGATCTGCAAGCCGACGACGACTGAACGAGA 227

17 uAsnLeuArgIleAlaLeuArgTyr 25
228 GAACCTCGGATCGGCTCGCTAC 252

seq_name: gb_pr4:AF022163

seq_documentation_block:
LOCUS AF022163 271 bp DNA PRI 05-JAN-1999
DEFINITION Homo sapiens isolate 138 MHC class I antigen HLA-H (HLA-H)
pseudogene, partial sequence.
ACCESSION AF022163
VERSION AF022163.1 GI:2655066
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE HLA-H: a pseudogene with increased variation due to balancing
JOURNAL Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE 99083426

```



```

/feature="HLA-H"
/pseudo
<1..>270
/number=2
<1..>270
/feature="HLA-H"
/feature="MHC class I antigen HLA-H"
/codon_start=1
/pseudo
<1..>271
/feature="HLA-H"
/pseudo
53 a 83 c 91 g 35 t 9 others
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 120.00      Length: 25
  Ratio: 4.800        Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 88.000

alignment_block:
US-08-653-294-32 x AF022165 ..
Align seg 1/1 to: AF022165 from: 1 to: 271
1 TTPASARGGLTGRGInilecysLysalaGlnThrAspArgG1 17
178 TGGACCGGACACACACAGATCTGCAAGGCCAGGCACAGACTGACACGAGA 227
17 uasnLeuArgileAlaLeuArgTyr 25
228 GAACCTGGGATCGCGCTCCGCTAC 252
seq_name: gb_pr4:AF022166

seq_documentation_block:
LOCUS AF022166 271 bp DNA PRI 05-JAN-1999
DEFINITION Homo sapiens isolate 219 MHC class I antigen HLA-H (HLA-H)
pseudogene, partial sequence.
ACCESSION AF022166
VERSION AF022166.1 GI:2655069
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE HLA-H: a pseudogene with increased variation due to balancing
selection at neighboring loci
JOURNAL Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE 99083426
REFERENCE
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1997) Fred Hutchinson Cancer Research Center,
1100 Fairview Ave. N., M374, Seattle, WA 98109, USA
FEATURES
Source
1..271
/organism="Homo sapiens"
/isolate="219"
/db_xref="taxon:9606"
/chromosomes="6"
/map="6p21.3"
/notes="African-American individual"
1..270
/feature="HLA-H"
/pseudo
/number=2
<1..>270
/feature="HLA-H"
/feature="MHC class I antigen HLA-H"
/codon_start=1
CDS

```

```

/feature="HLA-H"
/pseudo
<1..>271
/feature="HLA-H"
/pseudo
58 a 84 c 91 g 38 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 120.00      Length: 25
  Ratio: 4.800        Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 88.000

alignment_block:
US-08-653-294-32 x AF022166 ..
Align seg 1/1 to: AF022166 from: 1 to: 271
1 TTPASARGGLTGRGInilecysLysalaGlnThrAspArgG1 17
178 TGGACCGGACACACACAGATCTGCAAGGCCAGGCACAGACTGACACGAGA 227
17 uasnLeuArgileAlaLeuArgTyr 25
228 GAACCTGGGATCGCGCTCCGCTAC 252
seq_name: gb_pr4:AF022172

seq_documentation_block:
LOCUS AF022172 271 bp DNA PRI 05-JAN-1999
DEFINITION Gorilla gorilla isolate Shamba MHC class I antigen HLA-H ortholog
(HLA-H) gene, partial cds.
ACCESSION AF022172
VERSION AF022172.1 GI:2655079
KEYWORDS gorilla.
SOURCE Gorilla gorilla
ORGANISM Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE HLA-H: a pseudogene with increased variation due to balancing
selection at neighboring loci
JOURNAL Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE 99083426
REFERENCE
AUTHORS Grimsley, C., Mather, K.A. and Ober, C.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1997) Fred Hutchinson Cancer Research Center,
1100 Fairview Ave. N., M374, Seattle, WA 98109, USA
FEATURES
Source
1..271
/organism="Gorilla gorilla"
/isolate="Shamba"
/db_xref="taxon:9593"
<1..>271
/feature="HLA-H"
/number=2
<1..>270
/feature="HLA-H"
/codon_start=1
/product="MHC class I antigen HLA-H ortholog"
/protein_id="AAC99794.1"
/db_xref="GI:2655080"
/translation="GSHSMRYFTTMSRPGEPREISVGYVDDTQFVRFDSDASPR
EEPRAPWMEGPEYWDNRNTQICKAQATERNLRLALRYNQSEG"
59 a 84 c 91 g 37 t
BASE COUNT
ORIGIN

alignment_scores:

```

Quality: 120.00 Length: 25  
 Ratio: 4.800 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 88.000

## alignment\_block:

US-08-653-294-32 x AF022172 ..

Align seg 1/1 to: AF022172 from: 1 to: 271

1 TrpAspArgGlnThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17  
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 178 TGGGACCGGAACACACAGATCTCAAGGCCACACAGACTGAACGAGA 227  
 17 uAsnLeuArgIleAlaLeuArgTyr 25  
 |||||:::|||||:::|||||:::|||||:::|||||  
 228 GAACCTCGGATCGCTCGCTAC 252

seq\_name: gb\_pr2:HUMHRCAR3

seq\_documentation\_block: 1096 bp DNA PRI 14-FEB-1996  
 LOCUS HUMHRCAR3  
 DEFINITION Human MHC HLA-AR gene, inactive antigen.  
 ACCESSION M32106  
 VERSION M32106.1 GI:187800  
 KEYWORDS major histocompatibility complex.  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1096)  
 AUTHORS Zemmour, J., Koller, B.H., Ennis, P.D., Geraghty, D.E., Lawlor, D.A.,  
 Orr, H.T. and Parham, P.  
 TITLE HLA-AR, an inactivated antigen-presenting locus related to HLA-A.  
 JOURNAL Implications for the evolution of the MHC  
 MEDLINE J. Immunol. 144 (9), 3619-3629 (1990)  
 COMMENT 90229758  
 Authorin copy of sequence [1] kindly submitted by J. Zemmour,  
 14-FEB-1990.

## FEATURES

source location/Qualifiers  
 1..1096  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="hematopoietic B-lymphoblastoid cell line  
 LCL721"  
 CDS 1..>1096  
 /note="MHC HLA-AR; putative"  
 /pseudo  
 /codon\_start=1  
 BASE COUNT 222 a 334 c 362 g 178 t  
 ORIGIN

## alignment\_scores:

Quality: 120.00 Length: 25  
 Ratio: 4.800 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 88.000

## alignment\_block:

US-08-653-294-32 x HUMHRCAR3 ..

Align seg 1/1 to: HUMHRCAR3 from: 1 to: 1096

1 TrpAspArgGlnThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17  
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 250 TGGGACCGGAACACACAGATCTCAAGGCCACACAGACTGAACGAGA 299  
 17 uAsnLeuArgIleAlaLeuArgTyr 25  
 |||||:::|||||:::|||||:::|||||:::|||||  
 300 GAACCTCGGATCGCTCGCTAC 324

OM of: US-08-653-294-32 to: N\_Geneseq\_36:\* out\_format : pfs  
 Date: Feb 8, 2000 7:31 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=frame+p2n.model -DEV=xlbp  
 -Q=/cgnl1/USPIO-spool/US08653294/runat\_04022000\_160701\_15807/app\_query.fasta.2  
 -DB=N\_Geneseq\_36 -QMT=fastap -SUFFIX=ring -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsm62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pt  
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
 -MAXLEN=100000 -USBR=US08653294 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
 -THREADS=1

## Search information block:

Query: US-08-653-294-32  
 Query length: 25  
 Database: N\_Geneseq\_36:\*  
 Database sequences: 311585  
 Database length: 125096042  
 Search time (sec): 873.190000

score_list:	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:N70935	+	118.00	355.37	3.4e-12	1026	Sequence encoding the human h
N_Geneseq_36:N70225	+	118.00	342.41	1.8e-11	3874	Sequence of genomic DNA encodi
N_Geneseq_36:V61639	+	118.00	337.28	3.5e-11	6553	HLA B27 consensus sequence. De
N_Geneseq_36:Q01822	+	101.00	300.20	4.0e-09	1086	Sequence encoding HLA-B*52 ant
N_Geneseq_36:Q05701	+	101.00	300.18	4.0e-09	1089	HLA-B*52 gene for production o
N_Geneseq_36:Q29167	+	100.00	310.57	1.1e-09	270	HLA-B*52 exon 2 alpha-1 domai
N_Geneseq_36:Q01834	+	95.00	280.93	4.8e-08	1086	Sequence encoding HLA-B*51 ant
N_Geneseq_36:Q05693	+	95.00	280.90	4.8e-08	1089	HLA-B*51 gene for production o
N_Geneseq_36:Q12114	+	95.00	280.90	4.8e-08	1089	HLA-B*53 exon. HLA-B*53 gene,
N_Geneseq_36:Q75974	+	84.00	230.73	2.3e-05	4059	pHLA-B7 expression vector. New
N_Geneseq_36:Q75974	+	84.00	230.73	2.3e-05	4059	pHLA-B7/beta-2 microglobulin e
N_Geneseq_36:Q12116	+	70.00	200.49	0.0014	1101	HLA-C exon Cb-1. HLA-C gene, I
N_Geneseq_36:Q02693	+	66.00	201.42	0.0013	268	Human Class I non-classical HLA
N_Geneseq_36:Q12113	+	66.00	187.74	0.0074	1089	HLA-B*35 exon. HLA-B*35 gene -
N_Geneseq_36:Q12117	+	65.00	184.42	0.0113	1101	HLA-C exon Cb-2. HLA-C gene, I
N_Geneseq_36:N80603	+	62.00	173.29	0.0473	1284	Probe F10 of Major Histocompat
N_Geneseq_36:T21345	+	52.00	156.13	0.4273	277	Human gene signature HUMG02707
N_Geneseq_36:V12763	+	51.50	139.22	3.73	1329	Metal-regulated transporter ge
N_Geneseq_36:V34403	+	49.00	132.88	8.43	1118	Murine macrophage inflammatory
N_Geneseq_36:V34181	+	48.00	135.77	5.82	598	Human secreted protein gene 28
N_Geneseq_36:T29389	+	48.00	132.70	8.62	819	Canine interferon-beta gene. Pr
N_Geneseq_36:V30457	+	48.00	128.32	15.13	1284	Chimeric HLA-A2.1/beta-2 micro
N_Geneseq_36:Q92069	+	48.00	124.31	25.27	1935	Carbonic anhydrase cDNA. Clone
N_Geneseq_36:Q92066	+	48.00	123.16	29.30	2178	Maize carbonic anhydrase. Clon
N_Geneseq_36:Q92068	+	48.00	123.11	29.50	2190	Carbonic anhydrase cDNA. Clone
N_Geneseq_36:T86375	+	48.00	93.88	1.3e+03	43804	Chicken embryo lethal orphan
N_Geneseq_36:X26690	+	48.00	93.88	1.3e+03	43804	Complete genome sequence of a
N_Geneseq_36:Q70152	+	47.00	115.84	74.93	3319	K39 gene of Leishmania chagasi
N_Geneseq_36:T42166	+	47.00	115.84	74.93	3319	Leishmania chagasi K39 antigen
N_Geneseq_36:V18977	+	47.00	114.72	86.48	3722	Non-B, non-C, non-G hepatitis
N_Geneseq_36:Q04524	+	47.00	113.85	96.74	4071	Alpha-1-6-glucanase gene. Alph
N_Geneseq_36:N60080	+	46.50	106.29	255.06	7494	Complete nucleotide sequence c
N_Geneseq_36:V04901	+	46.00	121.05	38.42	1400	Nucleotide sequence of a port
N_Geneseq_36:V04900	+	46.00	120.81	39.59	1434	Nucleotide sequence of a port
N_Geneseq_36:Q25977	+	46.00	119.76	45.34	1598	MH mutant porcine ryanodine re
N_Geneseq_36:Q51474	+	45.00	155.32	0.4737	30	Oligonucleotide probe with phot
N_Geneseq_36:V32024	+	45.00	109.92	160.00	3150	Human p115 Rho-guanine exchang
N_Geneseq_36:V73804	+	45.00	86.41	3.3e+03	35100	KSHV LTR DNA (nucleotides 70,
N_Geneseq_36:V19941	+	45.00	73.09	1.8e+04	137507	KSHV long unique coding regi
N_Geneseq_36:Q37052	+	44.50	108.83	184.04	2988	Expression vector pAP045. Regi
N_Geneseq_36:Q43661	+	44.50	104.71	312.13	4558	Acetobacter cdg2 operon. Polyr

N\_Geneseq\_36:X40673 + 44.00 128.96 13.93 322 ! Human secreted protein 5' ES  
 N\_Geneseq\_36:V01882 + 44.00 109.17 176.22 2448 ! Human nel-related protein t  
 N\_Geneseq\_36:V38678 + 44.00 108.40 194.51 2649 ! Mus musculus SOCS10 gene. S  
 N\_Geneseq\_36:V01883 + 44.00 106.56 246.19 3198 ! Human nel-related protein t

seq\_name: N\_Geneseq\_36:N70935

## seq\_documentation\_block:

ID N70935 standard; DNA; 1026 BP.  
 AC N70935;  
 DT 10-APR-1991 (first entry)  
 DE Sequence encoding the human histocompatibility antigen HLA B27.  
 KW Rheumatic disorder; genetic screening; diagnosis;  
 KW ankylosing spondylitis; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 1.1026  
 PN DE3542024-A.  
 PD 04-JUN-1987.  
 PF 28-NOV-1985; 542024.  
 PR 28-NOV-1985; DE-542024.  
 PR 21-DEC-1985; DE-545576.  
 PA (BEHW) BEHRINGWERKE AG.  
 PI Rietmuller G, Meo T, Weiss E, Szots H;  
 DR WPI: 87-157893/23.  
 DR P-PSDB; P70590.  
 FT DNA coding for antigen HLA B27 - and diagnostic reagents contg.  
 PS such DNA, antigen or antibody  
 PS Claim 2; Page 4; 5pp; German.  
 CC The DNA may be used as a hybridisation probe for detecting the HLA  
 CC B27 gene, e.g. for assessing susceptibility to rheumatic disorders  
 CC such as ankylosis spondylitis, or may be used to transform cells  
 CC for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27  
 CC antibody in human serum, or to produce mono- or polyclonal HLA B27  
 CC antibodies for use in immunoassay.  
 SQ Sequence 1026 BP; 213 A; 307 C; 344 G; 162 T;

## alignment\_scores:

Quality: 118.00 Length: 25  
 Ratio: 5.130 Gaps: 0  
 Percent Similarity: 92.000 Percent Identity: 88.000

## alignment\_block:

US-08-653-294-32 x N70935 ..

Align seg 1/1 to: N70935 from: 1 to: 1026

1 TTPSPARGLUTHRGlnIleCysLysAlaLysAlaGlnThrAspArg1 17  
 |||||  
 178 TGGACCGGGGACACAGATCTGCAAGGCCAGGCACACTGACCGAGA 227  
 |||||  
 17 uAsnLeuArgIleAlaLeuArgTyr 25  
 |||||  
 228 GGACCTCGGACCTGCTCCGTAC 252  
 |||||

seq\_name: N\_Geneseq\_36:N70225

## seq\_documentation\_block:

ID N70225 standard; DNA; 3874 BP.  
 AC N70225;  
 DT 03-APR-1991 (first entry)  
 DE Sequence of genomic DNA encoding human histocompatibility antigen  
 DE HLA-B\*27.  
 KW Ankylosing spondylitis; rheumatic disorder; diagnosis; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT intron 518..590  
 FT intron 720..989  
 FT intron /\*tag= b  
 FT intron 1090..1506  
 FT intron /\*tag= c

```

FT Intron 1932..2357
FT Intron /*tag= d
FT Intron 2450..2566
FT Intron /*tag= e
FT Intron 3009..3041
FT Intron /*tag= f
FT Intron 3148..3191
FT Intron /*tag= g
PN EP-226059-A.
PD 24-JUN-1987.
PF 21-NOV-1986; 116139.
PR 01-JAN-1985; DE-542024.
PR 21-DEC-1985; DE-545576.
PA (BEHW ) BEHRINGERWERKE AG.
PI Szots H, Weiss E, Dornier C, Lang M, Meo T, Riethmuller G;
DR WPI; 87-171469/25.
DR P-PSDB; P70155.
PT DNA coding for human histocompatibility antigen HLA-B 27 - useful
PT for diagnosis and antigen and antibody prodn.
PS Claim 1; p6; 13pp; German.
CC The DNA may be used to detect the HLA-B 27 gene (opt. mutated) in
CC human genetic material. The HLA-B 27 may be used to detect anti-HLA-
CC B 27 antibodies in human serum. The antibodies may be used to
CC determine HLA-B 27 levels in human serum, e.g. for diagnosis of
CC rheumatic disorders, esp. ankylosing spondylitis.
SQ Sequence 3874 BP; 751 A; 1094 C; 1171 G; 858 T;

alignment_scores:
  Quality: 118.00 Length: 25
  Ratio: 5.130 Gaps: 0
Percent Similarity: 92.000 Percent Identity: 88.000

alignment_block:
US-08-653-294-32 x N70225 ..
Align seg 1/1 to: N70225 from: 1 to: 3874

1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
896 TGGGACCGGGAGACACAGATCTGCAAGCCCAAGCAGACGACTGACCGAGA 945

17 uAsnLeuArgIleAlaLeuArgTyr 25
946 GGACCTGGGACCTGCTCCGCTAC 970

seq_name: N_Geneseq_36:T61639

seq_documentation_block:
ID T61639 standard; DNA; 6553 BP.
AC T61639.
DT 05-JUN-1997 (first entry)
DE HLA B27 consensus sequence.
KW HLA B27; seronegative spondylarthropathy; ankylosing spondylitis;
KW Reiter's syndrome; arthritis; acute anterior uveitis; diagnosis;
SS: ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT mrna 3968..6653
FT /*tag= a
FT /*note= "HLA-B27 3' flanking region, downstream of
FT 3' untranslated region"
FT mrna 4112..4556
FT /*tag= b
FT /*note= "3' flanking region diagnostic for genetic
FT predisposition to SNSA"
FT mrna 4270..4556
FT /*tag= b
FT /*note= "3' flanking region diagnostic for genetic
FT predisposition to SNSA"
FT misc_difference 4495
FT /*tag= d
FT /*note= "absence of cytosine at this site is

```

```

FT PN WO9709450-A1.
FT PD 13-MAR-1997.
FT PF 16-AUG-1996; U13256.
FT PR 01-SEP-1995; US-522942.
FT PA (CEDA-) CEDARS SINAI MEDICAL CENT.
FT PI Tyau DB;
FT DR WPI; 97-192924/17.
PT Detecting pre-disposition to seronegative spondylarthropathies -
PT from the absence of a C residue at a specific position in the
PT 3'-flanking region of the HLA B27 allele
PS Claim 1; Page 52-56; 68pp; English.
CC Genetic predisposition to seronegative spondylarthropathies (SNSA)
CC is detected by determining the absence of a cytosine nucleotide in
CC the 3' flanking region (see also T61647-48) of an HLA-B gene at a
CC position corresponding to nucleotide 4495 of the HLA-B27 consensus
CC sequence given in T61639. Probes and primers (see also T61640-46)
CC based on this region can be used in diagnostic assays to detect the
CC genetic predisposition to SNSA, and permit the distinction of B27+
CC individuals who are resistant to SNSA from B27+ normal individuals
CC who are susceptible (but as yet unaffected) to such diseases.
SQ Sequence 6553 BP; 1443 A; 1619 C; 2017 G; 1474 T;

alignment_scores:
  Quality: 118.00 Length: 25
  Ratio: 5.130 Gaps: 0
Percent Similarity: 92.000 Percent Identity: 88.000

alignment_block:
US-08-653-294-32 x T61639 ..
Align seg 1/1 to: T61639 from: 1 to: 6553

1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
1057 TGGGACCGGGAGACACAGATCTGCAAGCCCAAGCAGACGACTGACCGAGA 1106

17 uAsnLeuArgIleAlaLeuArgTyr 25
1107 GGACCTGGGACCTGCTCCGCTAC 1131

seq_name: N_Geneseq_36:Q01822

seq_documentation_block:
ID Q01822 standard; DNA; 1086 BP.
AC Q01822;
DT 19-MAY-1991 (first entry)
DE Sequence encoding HLA-BW52 antigen.
KW Probe; HLA class I DNA; immunogen; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..1086
FT /*tag= a
FT EP-354580-A.
FT PD 14-FEB-1990.
FT PF 10-AUG-1989.
FT PR 11-AUG-1988; JP-200758.
FT PA (OLYU) Olympus Optical Co., Ltd.
FT PI Kano K, Takiguchi;
FT DR WPI; 90-045289/07.
FT DR P-PSDB; R03142.
FT New DNA for class I human leucocyte antigens and derived probes and
FT transformed cells, useful for DNA typing, as immunogens etc.
PS Claim 2; pp11-12; 23pp; English.
CC The HLA class I DNA can be used as a source of probes for use in DNA
CC typing. Transformed cells, which are useful as immunogens, can be
CC obtained by introducing these DNAs into eucaryotic cells.
SQ Sequence 1086 BP; 223 A; 335 C; 358 G; 170 T;

alignment_scores:
  Quality: 101.00 Length: 25

```

Ratio: 4.391 Gaps: 0  
Percent Similarity: 92.000 Percent Identity: 80.000

## alignment\_block:

US-08-653-294-32 x Q01822 ..

Align seg 1/1 to: Q01822 from: 1 to: 1086

```

1 TrpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
|||||
249 TGGACCGGGAGACACAGATCTCCAGACCAACACACAGACTTACCGAGA 298
|||||
17 uAsnLeuArgIleAlaLeuArgTyr 25
|||||
299 GAACCTGGGATCGGCTCGCTAC 323

```

seq\_name: N\_Geneseq\_36:Q05701

## seq\_documentation\_block:

ID Q05701 standard; DNA; 1089 BP.  
AC Q05701.  
DT 03-JAN-1991 (first entry)  
DE HLA-Bw52 gene for production of monoclonal antibodies.  
KW Allotype specific monoclonal anti-HLA antibodies; hybridomas;  
KW transgenic animals; HLA-Bw52 gene; ss.  
FH Key Location/Qualifiers  
FT exon 1..73  
FT /\*tag= a  
FT /number=1  
FT exon 74..343  
FT /\*tag= b  
FT /number=2  
FT /\*note="alpha 1-domain"  
FT exon 344..619  
FT /\*tag= c  
FT /number=3  
FT /\*note="alpha 2-domain"  
FT intron 620..895  
FT /\*tag= d  
FT /number=4  
FT /\*note="alpha 3-domain"  
FT exon 896..1012  
FT /\*tag= e  
FT /number=5  
FT exon 1013..1042  
FT /\*tag= f  
FT /number=6  
FT exon 1043..1089  
FT /\*tag= g  
FT /number=7

EP-383183-A.

PD 22-AUG-1990.

PF 07-FEB-1990; 102424.

PR 08-FEB-1989; JP-029313.

PA (OLYU ) OLYMPUS OPTICAL KK.

PI Takiguchi M;

DR WPI: 90-255479/34.

PT Allotype specific monoclonal anti- HLA antibodies prodn. - using  
PT hybridomas derived from transgenic animals carrying HLA gene and  
PT immunised with HLA antigen of different allotype  
PS Disclosure: Fig 1 A-G; 20pp; English.  
CC The human HLA-Bw52 gene was introduced into mouse L cells and  
CC then these cells used to immunise one of the transgenic mice  
CC (See Q05693).  
CC The spleen lymphocytes were fused with myeloma cells (P3x63-Ag8.653).  
CC Hybridomas producing antibodies were selected.  
SQ Sequence 1089 BP; 223 A; 336 C; 359 G; 171 T;

## alignment\_scores:

Quality: 101.00 Length: 25  
Ratio: 4.391 Gaps: 0  
Percent Similarity: 92.000 Percent Identity: 80.000

## alignment\_block:

US-08-653-294-32 x Q05701 ..

Align seg 1/1 to: Q05701 from: 1 to: 1089

```

1 TrpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
|||||
250 TGGACCGGGAGACACAGATCTCCAGACCAACACACAGACTTACCGAGA 299
|||||
17 uAsnLeuArgIleAlaLeuArgTyr 25
|||||
300 GAACCTGGGATCGGCTCGCTAC 324

```

seq\_name: N\_Geneseq\_36:Q29167

## seq\_documentation\_block:

ID Q29167 standard; DNA; 270 BP.  
AC Q29167.  
DT 09-MAR-1993 (first entry)  
DE HLA-Bw 52 exon 2 alpha-1 domain.  
KW Human leukocyte antigen; transgenic; germ cells; somatic cells;  
KW expression; ss.  
PN J04091731-A.  
PD 25-MAR-1992.  
PF 03-AUG-1990; 207329.  
PR 03-AUG-1990; JP-207329.  
PA (OLYU ) OLYMPUS OPTICAL CO.  
DR WPI: 92-342893/42.  
PT Transgenic non-human mammalian HLA-Bw 52 gene - useful for  
PT analysis of expression of gene structure, and prodn. of  
PT mouse model of human disease  
PS Disclosure: Fig 1; 8pp; Japanese.  
CC The sequence shows the exon 2 alpha-1-domain of the human leukocyte  
CC antigen-Bw 52 gene. The complete gene may be introduced into non-  
CC human mammals, pref. rat or mouse, or their ancestors at the primary  
CC developmental biological step via transplantation into the zygote or  
CC embryo to generate transgenic non-human mammals incorporating the  
CC HLA-Bw 52 gene in both their germ cells and somatic cells. Transgenic  
CC non-human mammals contg. HLA-Bw 52 are useful for the analysis of  
CC expression of the gene, its structure, and prodn. of mouse models of  
CC human disease. See also Q29166-72.  
SQ Sequence 270 BP; 59 A; 88 C; 86 G; 37 T;

## alignment\_scores:

Quality: 100.00 Length: 25  
Ratio: 4.348 Gaps: 0  
Percent Similarity: 92.000 Percent Identity: 80.000

## alignment\_block:

US-08-653-294-32 x Q29167 ..

Align seg 1/1 to: Q29167 from: 1 to: 270

```

1 TrpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
|||||
177 TGGACCGGGAGACACAGATCTCCAGACCAACACACAGACTTACCGAGA 226
|||||
17 uAsnLeuArgIleAlaLeuArgTyr 25
|||||
227 GAACCTGGGATCGGCTCGCTAC 251

```

seq\_name: N\_Geneseq\_36:Q01834

## seq\_documentation\_block:

ID Q01834 standard; DNA; 1086 BP.  
AC Q01834.  
DT 19-MAR-1991 (first entry)  
DE Sequence encoding HLA-B51 antigen.  
KW Probe; HLA class I DNA; immunogen; ss.  
OS Homo sapiens.  
PN EP354580-A.

PD 14-FEB-1990.  
 PF 10-AUG-1989.  
 PR 11-AUG-1988; JP-200758.  
 PA (OLYU) OLYMPUS Optical Co., Ltd.  
 PI Kano K, Takiguchi;  
 DR WPI; 90-046289/07.  
 PT New DNA for class 1 human leucocyte antigens and derived probes and  
 PT transformed cells, useful for DNA typing, as immunogens etc.  
 PS Claim 1; Page 11; 23pp; English.  
 CC The HLA class I DNA can be used as a source of probes for use in DNA  
 CC typing. Transformed cells, which are useful as immunogens, can be  
 CC obtained by introducing these DNAs into eucaryotic cells.  
 SQ Sequence 1086 BP; 224 A; 334 C; 356 G; 172 T;  
 alignment\_scores:  
 Quality: 95.00 Length: 25  
 Ratio: 4.130 Gaps: 0  
 Percent Similarity: 92.000 Percent Identity: 76.000  
 alignment\_block:  
 US-08-653-294-32 x Q01834 ..  
 Align seg 1/1 to: Q01834 from: 1 to: 1086  
 1 TtpspargluThrGlnileCysLysAlaGlnThrAspArgG1 17  
 |||||.....  
 249 TGGGACCGGACACACAGATCTTCAAGACCAACACACAGACTTACCGAGA 298  
 |||||.....  
 17 uAsnLeuArgIleAlaLeuArgTyr 25  
 |||||.....  
 299 GAACCTGGCGATCGCGTCCGCTAC 323  
 |||||.....  
 seq\_name: N\_Geneseq\_36:Q05693

seq\_documentation\_block:  
 ID Q05693 standard; DNA; 1089 BP.  
 AC Q05693;  
 DT 03-JAN-1991 (first entry)  
 DE HLA-B51 gene for production of monoclonal antibodies.  
 KW Allotype specific monoclonal anti-HLA antibodies; hybridomas;  
 KW transgenic animals; HLA-B51 gene; ss.  
 FH Key Location/Qualifiers  
 FT exon 1..73  
 FT /\*tag= a  
 FT /number=1  
 FT exon 74..343  
 FT /\*tag= b  
 FT /number=2  
 FT exon 344..619  
 FT /\*tag= c  
 FT /number=3  
 FT intron 620..895  
 FT /\*tag= alpha 2-domain  
 FT /number=4  
 FT exon 896..1012  
 FT /\*tag= e  
 FT /number=5  
 FT exon 1013..1042  
 FT /\*tag= f  
 FT /number=6  
 FT exon 1043..1089  
 FT /\*tag= g  
 FT /number=7  
 EP-383183-A.  
 PD 22-AUG-1990.  
 PF 07-FEB-1990; 102424.  
 PR 08-FEB-1989; JP-029313.  
 PA (OLYU) OLYMPUS OPTICAL KK.  
 PI Takiguchi M;

DR WPI; 90-255479/34.  
 PT Allotype specific monoclonal anti- HLA antibodies prodn. - using  
 PT hybridomas derived from transgenic animals carrying HLA gene and  
 PT immunised with HLA antigen of different allele  
 PS Disclosure; Fig 1 A-G; 20pp; English.  
 CC The human HLA-B51 gene was injected into fertilised mouse eggs and  
 CC then these introduced into the uterus of a pseudo pregnant mouse.  
 CC The young were tested to ensure incorporation of the gene into the  
 CC chromosome, and one of them mated 3 times with a normal male to  
 CC produce 16 young, seven of which carried the HLA-B51 gene.  
 CC The transgenic offspring were immunised with HLA antigen.  
 CC The spleen lymphocytes were fused with myeloma cells. Hybridomas  
 CC producing antibodies were selected.  
 CC See also Q05701.  
 SQ Sequence 1089 BP; 224 A; 335 C; 357 G; 173 T;

alignment\_scores:  
 Quality: 95.00 Length: 25  
 Ratio: 4.130 Gaps: 0  
 Percent Similarity: 92.000 Percent Identity: 76.000  
 alignment\_block:  
 US-08-653-294-32 x Q05693 ..  
 Align seg 1/1 to: Q05693 from: 1 to: 1089

1 TtpspargluThrGlnileCysLysAlaGlnThrAspArgG1 17  
 |||||.....  
 250 TGGGACCGGACACACAGATCTTCAAGACCAACACACAGACTTACCGAGA 299  
 |||||.....  
 17 uAsnLeuArgIleAlaLeuArgTyr 25  
 |||||.....  
 300 GAACCTGGCGATCGCGTCCGCTAC 324  
 |||||.....  
 seq\_name: N\_Geneseq\_36:Q12114

seq\_documentation\_block:  
 ID Q12114 standard; DNA; 1089 BP.  
 AC Q12114;  
 DT 23-AUG-1991 (first entry)  
 DE HLA-Bw53 exon.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 1..1089  
 FT /\*tag= a  
 FT J03112487-A.  
 PN 14-MAY-1991.  
 PF 22-SEP-1989; 247697.  
 PR 22-SEP-1989; JP-247697.  
 PA (OLYU) OLYMPUS OPTICAL KK.  
 DR WPI; 91-182991/25.  
 DR P-PSDB; R12463.  
 PT HLA-Bw53 gene, DNA probe and transformant cells - used for  
 PT immunisation, identifying specificity of antiserum etc.  
 PS Claim 1; Page 1; 11pp; Japanese.  
 CC Probes comprising part of the sequence can be used to identify  
 CC Class I genes. The DNA can be expressed for immunisation of  
 CC animals and prodn. of monoclonal antibodies specific for the  
 CC HLA-Bw53 antigen. See also J03112485 and J03112486.  
 SQ Sequence 1089 BP; 222 A; 337 C; 356 G; 174 T;

alignment\_scores:  
 Quality: 95.00 Length: 25  
 Ratio: 4.130 Gaps: 0  
 Percent Similarity: 92.000 Percent Identity: 76.000  
 alignment\_block:  
 US-08-653-294-32 x Q12114 ..





```

FT      ribosomal subunits when located within a mRNA
FT      mol. It enables the translational start codon of
FT      the beta-2 microglobulin, downstream of the HLA-B7
FT      stop codon on this bicistronic mRNA to be
FT      recognised by the ribosome"
FT      2480. .2839
FT      /*tag= 1
FT      /note= "encodes beta-2 microglobulin; this cDNA is
FT      deriv. from chimpanzee (differs to the human
FT      cDNA by only 4 bases)"
FT      2840. .2846
FT      /*tag= j
FT      /note= "3' untranslated region of the beta-2
FT      microglobulin mRNA"
FT      2847. .2870
FT      /*tag= k
FT      /note= "synthetic linker"
FT      3'UTR . .3111
FT      /*tag= 1
FT      /note= "bovine growth hormone 3'UTR and transcriptional
FT      terminator; it starts at a blunt-ended BglI site
FT      within the 3'UTR of the mRNA coding sequence"
FT      2979. .2984
FT      /*tag= m
FT      misc_feature
FT      3112. .3151
FT      /*tag= n
FT      /note= "synthetic linker to facilitate cloning"
FT      complement (3151. .3967)
FT      /*tag= o
FT      /note= "kanamycin resistance gene open reading frame;
FT      the gene is taken from the transposable element
FT      Tn903"
FT      4014. .4965
FT      /*tag= p
FT      /note= "pBR322 backbone contg. bacterial origin of
FT      replication, it represents nucleotides 2244-3193"
FT      WO9429469-A.
FT      22-DEC-1994.
FT      PD 27-MAY-1994; U06069.
FT      PR 07-JUN-1993; US-074344.
FT      PA (UNMI ) UNIV MICHIGAN.
FT      PI (VICA-) VICAL INC.
FT      DR WPI; 95-036494/05.
FT      PI New vectors for gene therapy, partic for tumours - comprising
FT      genetic material encoding one or more cistron(s) which express
FT      immunogenic or therapeutic peptide(s)
FT      PS Claim 8; Page 41-42; 50pp; English.
FT      CC The pHLA-B7/beta-2 microglobulin plasmid expression vector, in addition
FT      CC to the kanamycin resistance gene, contains the plasmid DNA encoding of a
FT      CC heavy (human HLA-B7) and light (beta-2 microglobulin) proteins of a
FT      CC class I major histocompatibility complex (MHC) antigen. The plasmid is
FT      CC designed to express these two proteins via a bicistronic mRNA in
FT      CC eukaryotic cells. Initiation of transcription of the mRNA is dependent
FT      CC on a Rous Sarcoma Virus (RSV) promoter sequence deriv. from the 3' long
FT      CC terminal repeat. Termination of transcription is dependent upon the
FT      CC polyadenylation signal sequence deriv. from the bovine growth hormone
FT      CC gene. Eukaryotic cell translation of the heavy chain is regulated by the
FT      CC 5' cap-dependent protein start site. Translation of the light chain is
FT      CC controlled by the CITE. Finally the replication of the plasmid in
FT      CC bacterial cells is controlled by the presence of a bacterial origin of
FT      CC replication. The vector is used partic. for the treatment of neoplastic
FT      CC disease, eg. melanoma, and provides enhanced gene delivery and expression
FT      CC in vivo.
FT      SQ Sequence 4965 BP; 1171 A; 1293 C; 1338 G; 1163 T;

alignment_scores:
    Quality: 84.00      Length: 20
    Ratio: 4.421      Gaps: 0
Percent Similarity: 95.000 Percent Identity: 80.000

alignment_block:

```

```

US-08-653-294-32 x Q75973 ..
Align seg 1/1 to: Q75973 from: 1 to: 4965

1 TrpAspArgGluThrGlnIleCysLysAlaGlnThrAspArgG1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
784 TGGGACCGGAGACACACAGATCTACAAGGCCGACGACTGACCGAGA 833

17 uAsnLeuArg 20
|:|:|:|:|:|
834 GAGCCTGCGG 843

seq_name: N_Geneseq_36:Q12116
seq_documentation_block:
ID_ Q12116 standard; DNA; 1101 BP.
AC Q12116;
DT 29-AUG-1991 (first entry)
DE HLA-C exon Cb-1.
KW Human leukocyte antigen; probe; major histocompatibility complex;
KW MHC; class I; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1. .1101
FT CDS /*tag= a
FT J03112485-A.
FT 14-MAY-1991.
FT PD 22-SEP-1989; 247695.
FT PR 22-SEP-1989; JP-247695.
FT PA (OLYU ) OLYMPUS OPTICAL KK.
FT DR WPI; 91-182989/25.
FT P-PSDB; R12465.
FT HLA-C gene, DNA probe and transformant cells - for immunisation
FT of animals and monoclonal antibody development.
FT PS Claim 1; Page 1; 13pp; Japanese.
FT CC Probes comprising part of the sequence can be used to identify
FT CC Class I genes. The DNA can be expressed for immunisation of
FT CC animals and prodn. of monoclonal antibodies specific for the HLA-C
FT CC antigen. See also Q12117 (same patent) and J03112486 and J03112487.
FT SQ Sequence 1101 BP; 211 A; 337 C; 337 G; 176 T;

alignment_scores:
    Quality: 70.00      Length: 20
    Ratio: 4.375      Gaps: 0
Percent Similarity: 80.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-32 x Q12116 ..
Align seg 1/1 to: Q12116 from: 1 to: 1101

1 TrpAspArgGluThrGlnIleCysLysAlaGlnThrAspArgG1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 TGGGACCGGAGACACACAGATCTACAAGGCCGACGACTGACCGAGT 299

17 uAsnLeuArg 20
|:|:|:|:|:|
300 GAGCCTGCGG 309

seq_name: N_Geneseq_36:V02693
seq_documentation_block:
ID_ V02693 standard; DNA; 268 BP.
AC V02693;
DT 19-MAY-1998 (first entry)
DE Human Class I non-classical HLA gene exon 2 consensus DNA.
KW Human leukocyte antigen class I gene; non-classical HLA; allele testing;
KW donor; tissue matching; recipient; graft rejection; class typing; ds.
OS Homo sapiens.
FT WO9723645-A1.
FT PN 03-JUL-1997.
FT PF 04-JAN-1996; U00362.

```

PR 04-JAN-1996; WO-U00362.  
 PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
 PI Cereb N, Yang SY;  
 DR WPI: 97-351080/32.  
 PT DNA-based human leukocyte antigen class I gene typing method -  
 PT useful for tissue matching and prevention of graft versus host  
 PT disease  
 PS Disclosure: Fig 6; 89pp; English.  
 CC V02693 is a consensus sequence of the human leukocyte antigen (HLA) Class  
 CC I non-classical HLA genes exon 2, which is used in a novel method for  
 CC testing a tissue sample to determine the allelic type of a HLA class I  
 CC gene in the sample. The HLA Class I gene is selected from among HLA-A,  
 CC -B and -C genes. The method comprises of treating the tissue sample to  
 CC obtain nucleic acid polymers suitable for amplification then combining  
 CC these polymers with a first primer which hybridises with a portion of  
 CC intron 1 or intron 3 of the HLA Class I gene and a second primer which  
 CC hybridises with a different portion of the HLA Class I gene under  
 CC conditions suitable for amplification to obtain an amplified product.  
 CC The product is then evaluated to determine the allelic type of the  
 CC HLA-Class I gene. The method is useful for tissue matching HLA class I  
 CC antigens between donors and recipients and hence for preventing graft  
 CC versus host disease.  
 CC Sequence 268 BP; 52 A; 86 C; 92 G; 36 T;

alignment\_scores:  
 Quality: 66.00 Length: 17  
 Ratio: 4.714 Gaps: 0  
 Percent Similarity: 82.353 Percent Identity: 76.471  
 alignment\_block:  
 US-08-653-294-32 x V02693 ..  
 Align seg 1/1 to: V02693 from: 1 to: 268  
 1 TTPAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17  
 ||||| :|||||  
 177 TGGGACCGGGAGACACAGACTTCAAGGCCACACACAGACTGACCGAGA 226  
 17 u 17  
 227 G 227

seq\_name: N\_Geneseq\_36:Q12115  
 seq\_documentation\_block:  
 ID Q12115 standard; DNA; 1089 BP.  
 AC Q12115;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-B35 exon.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 1..1089  
 FT /\*tag= a  
 PN J03112486-A.  
 PD 14-MAY-1991.  
 PF 22-SEP-1989; 247697.  
 PR 22-SEP-1989; JP-247697.  
 PA (OLYU ) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182991/25.  
 DR P-PSDB; R12464.  
 PT HLA-B35 gene - used in DNA probe and transformant cells for  
 PT immunising animals, for developing monoclonal antibody.  
 PS Claim 1; Page 1; 11pp; Japanese.  
 CC Probes comprising part of the sequence can be used to identify  
 CC Class I genes. The DNA can be expressed for immunisation of  
 CC animals and prodn. of monoclonal antibodies specific for the  
 CC HLA-B35 antigen. See also J03112485 and J03112487.  
 CC Sequence 1089 BP; 221 A; 336 C; 359 G; 173 T;

alignment\_scores:  
 Quality: 66.00 Length: 20  
 Ratio: 3.667 Gaps: 0  
 Percent Similarity: 90.000 Percent Identity: 65.000  
 alignment\_block:  
 US-08-653-294-32 x Q12115 ..  
 Align seg 1/1 to: Q12115 from: 1 to: 1089  
 1 TTPAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17  
 ||||| :|||||  
 250 TGGGACCGGGAGACACAGACTTCAAGGCCACACACAGACTGACCGAGA 299  
 17 uAsnLeuArg 20  
 |||||  
 300 GAGCCTCGG 309  
 seq\_name: N\_Geneseq\_36:Q12117  
 seq\_documentation\_block:  
 ID Q12117 standard; DNA; 1101 BP.  
 AC Q12117;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-C exon Cb-2.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 1..1101  
 FT /\*tag= a  
 PN J03112485-A.  
 PD 14-MAY-1991.  
 PF 22-SEP-1989; 247695.  
 PR 22-SEP-1989; JP-247695.  
 PA (OLYU ) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182989/25.  
 DR P-PSDB; R12466.  
 PT HLA-C gene, DNA probe and transformant cells - for immunisation  
 PT of animals and monoclonal antibody development.  
 PS Claim 2; Page 1; 13pp; Japanese.  
 CC Probes comprising part of the sequence can be used to identify  
 CC Class I genes. The DNA can be expressed for immunisation of  
 CC animals and prodn. of monoclonal antibodies specific for the HLA-C  
 CC antigen. See also Q12116 (same patent) and J03112486 and J03112487.  
 CC Sequence 1101 BP; 215 A; 335 C; 379 G; 172 T;

alignment\_scores:  
 Quality: 65.00 Length: 20  
 Ratio: 4.062 Gaps: 0  
 Percent Similarity: 80.000 Percent Identity: 65.000  
 alignment\_block:  
 US-08-653-294-32 x Q12117 ..  
 Align seg 1/1 to: Q12117 from: 1 to: 1101  
 1 TTPAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17  
 ||||| :|||||  
 250 TGGGACCGGGAGACACAGAGTACAGCGCCAGCGCTGACCGAGT 299  
 17 uAsnLeuArg 20  
 |||||  
 300 GAGCCTCGG 309

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-32 to: EST.\* out\_format : pfs

Date: Feb 8, 2000 6:23 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlip  
-Q/cgml1/USFTO.spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.2  
-DB-EST -Qfmt=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELXT=7.000 -START=1 -MAPFIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-32  
Query length: 25  
Database: EST.\*  
Database sequences: 4538634  
Database length: 1887831982  
Search time (sec): 7600.090000

score_list:	Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_est38:AL036690	+	118.00	363.01	1.9e-11	171	!	AL036690 DKFZp564D2463.r1 564
gb_est38:D82221	+	106.00	318.34	5.8e-09	375	!	D82221 HUMHC4626 Human pancrea
gb_est6:D82189	+	100.00	298.74	7.1e-08	415	!	D82189 HUMHC4524 Human pancrea
gb_est23:AL124815	+	96.00	286.33	3.5e-07	414	!	AL124815 am56e06.x1 Johnston fi
gb_est10:AL151891	+	95.00	287.77	2.9e-07	255	!	AL151891 zc01f06.r1 Stratagene
gb_est11:AA263158	+	95.00	286.79	3.3e-07	283	!	AA263158 PW0534 KGI-a Lambda 2
gb_est26:AA1359260	+	95.00	279.46	8.4e-07	618	!	AA1359260 qv27b07.x1 NCI.CGAP.B
gb_est31:AA1696864	+	95.00	277.67	1.1e-06	748	!	AA1696864 WC74b11.x1 NCI.CGAP.P
gb_est6:D82177	+	94.00	282.24	5.9e-07	330	!	D82177 HUMHC4504 Human pancrea
gb_est10:AA158109	+	91.00	269.11	3.2e-06	495	!	AA158109 zc57h10.s1 Stratagene
gb_est10:AA147151	+	91.00	267.60	3.9e-06	581	!	AA147151 zc32d06.r1 Stratagene
gb_est12:R13904	+	88.00	260.49	9.6e-06	459	!	R13904 yf62c03.r1 Soares infant
gb_est9:AA121088	+	85.00	252.20	2.8e-05	411	!	AA121088 zm22d05.r1 Stratagene
gb_est10:AA143626	+	84.00	250.93	3.3e-05	338	!	AA143626 zc65h01.r1 Stratagene
gb_est11:T52124	+	81.00	240.83	0.0001	367	!	T52124 yb29g06.r1 Stratagene fe
gb_est12:AA310808	+	80.00	235.95	0.0002	443	!	AA310808 EST181593 Jurkat T-cell
gb_est9:AA112349	+	79.00	234.38	0.0003	376	!	AA112349 zm27b04.r1 Stratagene
gb_est2:R20954	+	78.00	227.42	0.0007	567	!	R20954 yg06c01.r1 Soares infant
gb_est14:AA263135	+	75.00	215.98	0.0029	710	!	AA263135 PW0598 KGI-a Lambda 2
gb_est27:F13067	+	74.00	221.08	0.0015	296	!	F13067 HSC31C011 normalized inf
gb_est27:AA1466429	+	74.00	219.09	0.0019	366	!	AA1466429 vx35B04.y1 Stratagene
gb_est8:C03945	+	72.00	217.15	0.0025	232	!	C03945 C03945 Human heart cDNA
gb_est13:AA352960	+	72.00	215.73	0.0030	270	!	AA352960 EST61101 Activated T-c
gb_est11:T10924	+	72.00	214.77	0.0034	299	!	T10924 hbc122 Human pancreatic
gb_est35:AA1826037	+	72.00	212.47	0.0045	382	!	AA1826037 wk14g03.x1 NCI.CGAP.Ly
gb_est17:AA663896	+	72.00	210.35	0.0060	479	!	AA663896 ae74d01.s1 Stratagene
gb_est20:AA879637	+	71.00	210.27	0.0060	347	!	AA879637 vx38B05.r1 Stratagene
gb_est9:AA100848	+	71.00	206.80	0.0094	502	!	AA100848 zm26c05.r1 Stratagene
gb_est13:AA361489	+	70.00	209.44	0.0067	272	!	AA361489 EST71032 T-cell lympho
gb_est14:AA449607	+	70.00	206.81	0.0094	360	!	AA449607 zx06h02.r1 Soares tota
gb_est21:AA970332	+	70.00	205.28	0.0114	424	!	AA970332 op21e09.s1 NCI.CGAP.Cc
gb_est10:AA151951	+	70.00	204.49	0.0126	461	!	AA151951 zc30a04.r1 Stratagene
gb_est14:AA401643	+	70.00	203.66	0.0141	504	!	AA401643 zc67c02.r1 Soares test
gb_est9:CI9056	+	69.50	209.02	0.0071	241	!	CI9056 C19056 Human placenta ch
gb_est13:AA381090	+	68.00	202.09	0.0172	307	!	AA381090 EST94155 Activated T-c
gb_est7:W52004	+	68.00	200.39	0.0214	368	!	W52004 zc92e11.r1 Pancreatic Is
gb_est10:AA127128	+	67.00	195.46	0.0403	447	!	AA127128 z187d05.r1 Stratagene
gb_est35:AL039796	+	67.00	194.35	0.0464	503	!	AL039796 DKFZp43B192.r1 434
gb_est44:AA208428	+	67.00	191.33	0.0684	594	!	AA208428 uc06c03.x1 NCI.CGAP.Me
gb_est13:AA352534	+	66.00	193.75	0.0232	203	!	AA352534 EST60562 Activated T-c
gb_est10:AA160317	+	66.00	193.27	0.0533	405	!	AA160317 zc56c07.r1 Stratagene
gb_est7:W40489	+	66.00	192.78	0.0568	427	!	W40489 zc48b01.r1 Pancreatic Is

gb\_est11:AA227277 + 65.00 190.89 0.0724 375 ! AA227277 z120h03.r1 Stratage  
gb\_est4:H23327 + 65.00 188.59 0.0972 479 ! H23327 ym52e08.r1 Soares inf  
gb\_est44:AW215398 + 65.00 186.62 0.1251 591 ! AW215398 up03c01.x1 NCI.CGAP

seq\_name: gb\_est38:AL036690

seq\_documentation\_block:  
LOCUS AL036690 171 bp mRNA EST 27-SEP-1999  
DEFINITION DKFZp564D2463.r1 564 (synonym: hfbz2) Homo sapiens cDNA clone  
DRFZp564D2463 5', mRNA sequence.

ACCESSION AL036690  
VERSION AL036690.3 GI:5927859  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 171)  
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and  
Wiemann,S.  
TITLE EST (Duesterhoeft, et al.)  
JOURNAL Unpublished (1999)  
COMMENT On Jul 7, 1999 this sequence version replaced gi:5866258.  
Contact: Duesterhoeft A  
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by Qiagen within the cDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
source  
1..171  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp564D2463"  
/tissue\_type="brain"  
/dev\_stage="fetal"  
/lab\_host="xl-2blue"  
/note="vector: pAMP1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 36 a 53 c 60 g 22 t  
ORIGIN

alignment\_scores:  
Quality: 118.00 Length: 25  
Ratio: 5.130 Gaps: 0  
Percent Similarity: 92.000 Percent Identity: 88.000

alignment\_block:  
US-08-653-294-32 x AL036690 ..

Align seg 1/1 to: AL036690 from: 1 to: 171  
1 TTPAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17  
|||||  
17 TGGGACGGGAGACACAGATCTGCAAGGCAAGGACACTGACGAGA 66  
17 uAsnLeuArgIleAlaLeuArgTyr 25  
|||||  
67 GGACCTCGGACCCCTGCTCCGTAC 91

seq\_name: gb\_est6:D82221

seq\_documentation\_block:  
LOCUS D82221 375 bp mRNA EST 09-FEB-1996  
DEFINITION HUMHB4626 Human pancreatic islet Homo sapiens cDNA similar to  
HLA-B, mRNA sequence.

```

ACCESSION      D82221
VERSION        D82221.1
KEYWORDS       EST
SOURCE         human
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 375)
AUTHORS        Takeda,J.
TITLE          Human pancreatic islet ESTs
JOURNAL        Unpublished (1995)
COMMENT        On Apr 14, 1993 this sequence version replaced gi:785255.
               Institute for Molecular and Cellular Regulation, Gunma University
               3-39-15 Showa-machi, Maebashi Gunma 371, Japan
               Tel: 272-20-8856
               Fax: 272-20-8896
               Email: jtakeda@sb.gunma-u.ac.jp.
FEATURES       Location/Qualifiers
               1..375
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="Human pancreatic islet"
               /note="vector: lambda ZAPII; Site_1: Eco RI; Site_2: Xho
               I; mRNA was prepared from normal adult human islets. cDNA
               was directionally synthesized from the Xho I in the vector
               to the EcoRI site. cDNA was size fractionated to remove
               sequences <1000 bp in size."
               75 a 124 c 118 g 55 t 3 others
               BASE COUNT
               ORIGIN
alignment_scores
  Quality: 106.00      Length: 25
  Ratio: 4.417        Gaps: 0
  Percent Similarity: 96.000      Percent Identity: 80.000
alignment_block
  US-08-653-294-32 x D82221 ..
  Align seg 1/1 to: D82221 from: 1 to: 375
1 TrpAspArgIuThrGlnIleCysLysAlaLysAlaGlnThrAspArgI 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
261 TGGACCGGAACACACAGATCTGCAAGACCAACACACAGACTTACCGAGA 310
17 uAsnLeuArgIleAlaLeuArgTyr 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
311 GAACCTCGGATCGCTCCGNTAC 335
seq_name: gb_est6:D82189
seq_documentation_block:
LOCUS      D82189          415 bp      mRNA      EST      09-FEB-1996
DEFINITION HUMHBC4524 Human pancreatic islet Homo sapiens cDNA similar to
            HLA-B, mRNA sequence.
ACCESSION  D82189
VERSION    D82189.1
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 415)
            Takeda,J.
            Human pancreatic islet ESTs
            Unpublished (1995)
            On Apr 14, 1993 this sequence version replaced gi:785206.
            Contact: Jun Takeda
            Institute for Molecular and Cellular Regulation, Gunma University
            3-39-15 Showa-machi, Maebashi Gunma 371, Japan
            Tel: 272-20-8856
            Fax: 272-20-8896

```

```

/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: Bluescript SK-; Site_1:
EcoRI; Stanley Neuropathology Consortium
(www.stanleylab.org) brains S-58, S-65,
Random + oligo-dT primed into EcoRI site of ZAP II Vector.
Mass excised. Avg insert length 1.9kb. Custom library
provided by Dr. Nancy Johnston [(410) 614-3918,
nlj@welchlink.welch.jhu.edu].
BASE COUNT      80 a 140 c 136 g 58 t
ORIGIN

```

```

alignment_scores:
  Quality: 96.00      Length: 25
  Ratio: 4.364       Gaps: 0
  Percent Similarity: 88.000 Percent Identity: 76.000

```

```
alignment_block:

```

```
US-08-653-294-32 x A1124815 ..
```

```
Align seg 1/1 to: A1124815 from: 1 to: 414
```

```

1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
239 TGGGACCGGAGACACAGATCTCCAGACCACACACAGACTTACCGAGA 288

```

```

17 uAsnLeuArgIleAlaLeuArgTyr 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
289 GAACCTGCGCAGCGCGCTCGGTAC 313

```

```
seq_name: gb_est10:A1151891
```

```
seq_documentation_block:
```

```

LOCUS      A1151891      255 bp      mRNA      EST      10-DEC-1996
DEFINITION zool106.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
IMAGE:566435 5' similar to gb:M15497.cds1 HLA CLASS I
HISTOCOMPATIBILITY ANTIGEN, A-24(A-9) A*2401 (HUMAN);, mRNA
sequence.

```

```

ACCESSION  A1151891
VERSION    A1151891.1 GI:1720754
KEYWORDS   EST.
SOURCE     human.

```

```

ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE  1 (bases 1 to 255)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chisnoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Scheinberg,K., Soares,M.B., Tan,F., Thierry-Weg,J.,
            Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
            and Marra,M.

```

```

TITLE      Generation and analysis of 280,000 human expressed sequence tags
JOURNAL    Genome Res. 6 (9), 807-828 (1996)
MEDLINE    9704478
COMMENT    On May 8, 1995 this sequence version replaced gi:800234.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810

```

```

Email: est@watson.wustl.edu

```

```

This clone is available royalty-free through LINDL ; contact the
IMAGE Consortium (info@image.lindl.gov) for further information.
Trace considered overall poor quality
Seq primer: ~28M13 rev2 from Amersham
High quality sequence stop: 1.
            Location/Qualifiers
              1..255
                /organism="Homo sapiens"
                /db_xref="GDB:459088"

```

```
FEATURES
source

```

```

BASE COUNT      57 a 70 c 75 g 44 t
ORIGIN

```

```

alignment_scores:
  Quality: 95.00      Length: 25
  Ratio: 4.524       Gaps: 0
  Percent Similarity: 84.000 Percent Identity: 76.000

```

```
alignment_block:

```

```
US-08-653-294-32 x AA151891 ..
```

```
Align seg 1/1 to: AA151891 from: 1 to: 255
```

```

1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
32 TGGGACGAGGAGACAGGGAAGTGAAGGCCCTCACAGACTGACCGCAGA 81

```

```

17 uAsnLeuArgIleAlaLeuArgTyr 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
82 GAACCTGCGGATCGCNCCTCGGTAC 106

```

```
seq_name: gb_est11:AA263158
```

```
seq_documentation_block:
```

```

LOCUS      AA263158      283 bp      mRNA      EST      02-JUL-1998
DEFINITION PMY0534 KGI-a Lambda zap Express cDNA library Homo sapiens cDNA 5',
IMAGE:566435 5' similar to gb:M15497.cds1 HLA CLASS I
HISTOCOMPATIBILITY ANTIGEN, A-24(A-9) A*2401 (HUMAN);, mRNA
sequence.

```

```

ACCESSION  AA263158
VERSION    AA263158.1 GI:1898964
KEYWORDS   EST.
SOURCE     human.

```

```

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE  1 (bases 1 to 283)
AUTHORS   Claudio,J.O., Liew,C.C., Dempsey,A.A., Cukerman,E., Stewart,A.K.,
            Na,E., Atkins,H.I., Iscove,N.N. and Hawley,R.G.
            Identification of sequence-tagged transcripts differentially
            expressed within the human hematopoietic hierarchy
            Genomics 50 (1), 44-52 (1998)

```

```

JOURNAL    98292493
MEDLINE

```

```

COMMENT    On May 5, 1995 this sequence version replaced gi:797810.
            Contact: Hawley RG
            Oncology Research Laboratories
            The Toronto Hospital
            CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
            Tel: 416 3403834
            Fax: 416 3403453
            Email: r.hawley@utoronto.ca
            Similar to M58636 MHC class I HLA-Bw gene. Clone was randomly
            picked from KGIa primary library.
            Seq primer: 5' GAAATTAACCTCACTAAGGG 3'
            High quality sequence stop: 283.
            Location/Qualifiers
              1..283
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="KGI-a Lambda Zap Express cDNA library"
                /cell_type="promyeloblast"
                /cell_line="KGI-a"
                /note="Vector: Lambda Zap Express (Stratagene); Site_1:
                EcoRI; Site_2: XhoI; Unidirectional cloning sites:

```





```

alignment_scores:
  Quality: 95.00      Length: 25
  Ratio: 4.130        Gaps: 0
  Percent Similarity: 92.000      Percent Identity: 76.000

alignment_block:
  US-08-653-294-32 x AI696864 ..
  Align seg 1/1 to: AI696864 from: 1 to: 748
    1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
    92 TGGACCGGAAACACACAGATCTTCAGACCAACACACAGACTTACCGAGA 141
    17 uAsnLeuArgIleAlaLeuArgTyr 25
    142 GAACCTGCGGATCGCGCTCGCTAC 166

seq_name: gb_est6:D82177

seq_documentation_block:
  LOCUS D82177 330 bp mRNA EST 09-FEB-1996
  DEFINITION HUMHBC4504 Human pancreatic islet Homo sapiens cDNA similar to
  HLA-B, mRNA sequence.
  ACCESSION D82177
  VERSION D82177.1 GI:1183645
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 330)
  AUTHORS Takeda,J.
  TITLE Human pancreatic islet ESTs
  JOURNAL Unpublished (1995)
  COMMENT On Apr 14, 1993 this sequence version replaced gi:785191.
  Contact: Jun Takeda
  Institute for Molecular and Cellular Regulation, Gunma University
  3-39-15 Showa-machi, Maebashi Gunma 371, Japan
  Tel: 272-20-8856
  Fax: 272-20-8896
  Email: jtakeda@sb.gunma-u.ac.jp.
  FEATURES
    source
      1..330
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="Human pancreatic islet"
      /note="Vector: Lambda ZAPII; Site.1: Eco RI; Site.2: Xho
      I; mRNA was prepared from normal adult human islets. cDNA
      was directionally synthesized from the Xho I in the vector
      to the EcoRI site. cDNA was size fractionated to remove
      sequences <1000 bp in size."
      62 a 109 c 106 g 49 t 4 others
  BASE COUNT
  ORIGIN
    alignment_scores:
      Quality: 94.00      Length: 23
      Ratio: 4.273        Gaps: 0
      Percent Similarity: 95.652      Percent Identity: 78.261

alignment_block:
  US-08-653-294-32 x D82177 ..
  Align seg 1/1 to: D82177 from: 1 to: 330
    1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
    261 TGGGACCGGAAACACACAGATCTTCAGACCAACACACAGACTTACCGAGA 310
    17 uAsnLeuArgIleAlaLeu 23

```

```

|||||
311 GAACCTGCGGATCGCGCTC 329

seq_name: gb_est10:AA158109

seq_documentation_block:
  LOCUS AA158109 495 bp mRNA EST 09-MAR-1998
  DEFINITION z057h10.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
  IMAGE:591043 3' similar to gb:U04245 HLA CLASS I HISTOCOMPATIBILITY
  ANTIGEN, B*7 B*0702 ALPHA (HUMAN);, mRNA sequence.
  ACCESSION AA158109
  VERSION AA158109.1 GI:1732894
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 495)
  AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
  Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
  Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
  Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  WashU-NCI human EST Project
  Unpublished (1997)
  COMMENT On Sep 12, 1996 this sequence version replaced gi:1395000.
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@wustl.edu
  This clone is available royalty-free through LNL: contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Possible reversed clone: polyt not found
  Insert Length: 1581 Std Error: 0.00
  Seq primer: -40M13 fwd. from Amersham
  High quality sequence stop: 43.
  FEATURES
    source
      1..495
      /organism="Homo sapiens"
      /db_xref="GDB:4622429"
      /db_xref="taxon:9606"
      /clone="IMAGE:591043"
      /clone_lib="Stratagene pancreas (#937208)"
      /lab_host="SOLR cells (kanamycin resistant)"
      /note="Organ: pancreas; Vector: pBluescript SK-; Site.1:
      EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:
      Oligo dt. Pancreatic adenocarcinoma cell line. Average
      insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
      sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
      CTCGAGTGTGTTTTTTTTTTT 3"
      95 a 156 c 163 g 75 t 6 others
  BASE COUNT
  ORIGIN
    alignment_scores:
      Quality: 91.00      Length: 20
      Ratio: 4.789        Gaps: 0
      Percent Similarity: 95.000      Percent Identity: 85.000

alignment_block:
  US-08-653-294-32 x AA158109 ..
  Align seg 1/1 to: AA158109 from: 1 to: 495
    1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
    224 TGGGACCGGAAACACACAGATCTGCAAGGCCAGGCACAGACTGACCGAGT 273
    17 uAsnLeuArg 20
    274 GGSCCTGCGG 283

```



|||||  
291 GAACCTGGGATCGCGCTCCG 312

seq\_name: gb\_est9:AA121088

seq\_documentation\_block:

LOCUS AA121088 411 bp mRNA EST 23-DEC-1997  
DEFINITION zM22d05.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone  
IMAGE:526377 5' similar to gb:M24034.cdsl HLA CLASS I  
HISTOCOMPATIBILITY ANTIGEN, BW-42 B\*4201 ALPHA (HUMAN);, mRNA  
sequence.

ACCESSION AA121088  
VERSION AA121088.1 GI:1678632  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 411)

REFERENCE Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 9704478

COMMENT On May 18, 1995 this sequence version replaced gi:811358.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent  
plate of this clone contains both human and mouse derived clones.  
Thus, the origin of this clone is uncertain. This caution should be  
kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham.

FEATURES  
source

1..411  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:3917806"  
/db\_xref="taxon:9606"  
/clone="IMAGE:526377"

/clone\_lib="Stratagene pancreas (#937208)"

/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pancreatic adenocarcinoma cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor  
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTT 3'."

BASE COUNT 75 a 133 c 136 g 62 t 5 others

ORIGIN

alignment\_scores:

Quality: 85.00 Length: 20  
Ratio: 4.722 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 80.000

alignment\_block:

US-08-653-294-32 x AA121088 ..

Align seg 1/1 to: AA121088 from: 1 to: 411

1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17

|||||  
236 TGGGACCGGACACANAGATCTGCAGGCCAGGCACACTGACCGACT 285

17 uAsnLeuArg 20

|||||

286 GGGCCTGCGG 295

seq\_name: gb\_est10:AA143626

seq\_documentation\_block:

LOCUS AA143626 338 bp mRNA EST 08-NOV-1997  
DEFINITION z065h01.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone  
IMAGE:591793 5' similar to gb:U04245 HLA CLASS I HISTOCOMPATIBILITY  
ANTIGEN, B-7 B\*0702 ALPHA (HUMAN);, mRNA sequence.

ACCESSION AA143626  
VERSION AA143626.1 GI:1713058  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 338)

REFERENCE

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

TITLE Unpublished (1997)

JOURNAL

COMMENT

On Nov 29, 1993 this sequence version replaced gi:634605.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1677 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 193.

FEATURES

source

1..338  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:4623179"  
/db\_xref="taxon:9606"  
/clone="IMAGE:591793"

/clone\_lib="Stratagene pancreas (#937208)"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pancreatic adenocarcinoma cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor  
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTT 3'."

BASE COUNT 64 a 97 c 113 g 54 t 10 others

ORIGIN

alignment\_scores:

Quality: 84.00 Length: 16  
Ratio: 5.250 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 93.750

alignment\_block:

US-08-653-294-32 x AA143626 ..

Align seg 1/1 to: AA143626 from: 1 to: 338

1 TtpAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArg 16

|||||

187 TGGGACCGGACACACAGATCTGCAAGGCCAGGCACACTGACCGA 234

seq\_name: gb\_est1:T52124

```

seq_documentation_block: 367 bp mRNA EST 06-FEB-1995
LOCUS T52124
DEFINITION Y29g06.r1 Stratagene fetal spleen (#937205) Homo sapiens cDNA
clone IMAGE:72634 5' similar to gb:M24032_cds1 HLA CLASS
I HISTOCOMPATIBILITY ANTIGEN, BW-65(B-14) B*1402 (HUMAN), mRNA
sequence.
ACCESSION T52124
VERSION T52124.1 GI:653984
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 367)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,E., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 48
High quality sequence stops: 330 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Seq primer: M13Rpi
High quality sequence stop: 330.
FEATURES
source
1..367
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:494299"
/db_xref="taxon:9606"
/clone="IMAGE:72634"
/clone_lib="Stratagene fetal spleen (#937205)"
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: spleen; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled spleens. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"
BASE COUNT 69 a 119 c 116 g 59 t 4 others
ORIGIN

alignment_scores:
Quality: 81.00 Length: 20
Ratio: 4.263 Gaps: 0
Percent Similarity: 95.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-32 x T52124
Align seg 1/1 to: T52124 from: 1 to: 367
1 TipAspArgGluThrGlnIleCysLysAlaLysAlaGlnThrAspArgG1 17
|||||
236 TGGGACCGGACACACAGTCTGCAAGACACACACAGACTGACCGAGA 285
17 uasnLeuArg 20
|:::|||||
286 GAGCTGCGG 295

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 04:05:44 ; Search time 133.56 Seconds  
(without alignments)  
2.660 Million cell updates/sec

Title: US-08-653-294-33

Perfect score: 74

Sequence: 1 KAQTDRENLRALRY 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	100.0	15	R92912	HLA-B*2702 CTL modu
2	74	100.0	15	W33795	Peptide B2702.70-8
3	74	100.0	25	R41205	Peptide fragment o
4	74	100.0	25	R83090	HLA-B*2702 CTL modu
5	74	100.0	25	R95416	HLA-B*2702.60-84. C
6	74	100.0	25	W33794	Peptide B2702.60-8
7	59	79.7	17	R71442	Human HLA-B*27-(62-
8	59	79.7	25	R41221	Peptide fragment o
9	59	79.7	25	R83091	HLA-B*2702 CTL modu
10	59	79.7	25	R95417	HLA-B*2705.60-84. C
11	59	79.7	337	P70590	Sequence of the hu
12	59	79.7	362	P70155	Sequence encoded b
13	56	75.7	25	R42886	Peptide fragment o
14	56	75.7	25	R83093	HLA-B*2702 CTL modu
15	56	75.7	25	R95422	HLA-B*2702.60-84. C
16	56	75.7	184	Y06801	Peptide seq ID No:
17	56	75.7	362	R03142	Sequence of HLA-Bw
18	56	75.7	362	R03144	Sequence of HLA-B5
19	56	75.7	362	R12463	HLA-B*53 exon. HLA
20	53	71.6	26	R83071	CTL modulating pep
21	52	70.3	16	R50266	HLA B27 hypervaria
22	50	67.6	17	R71443	Human [Phe74]-HLA-
23	50	67.6	21	W58992	Human HLA-B*27 alph
24	49	66.2	10	R41208	Peptide fragment o
25	49	66.2	10	R83062	HLA-B*2702 CTL modu
26	49	66.2	10	R95413	Alpha-helix of HL
27	49	66.2	10	W07512	T-cell modulating
28	49	66.2	10	W47265	Immunomodulatory p
29	49	66.2	10	W33784	Peptide B2702.75-8
30	49	66.2	20	R92907	HLA-B*2702 CTL modu
31	49	66.2	20	R92908	HLA-B*2702 CTL modu
32	49	66.2	20	R95428	HLA-B*2702 84-75-84
33	49	66.2	20	W33778	Immunomodulating d
34	49	66.2	20	W33791	Peptide B2702.84-7

35 45 60.8 25 1 R95431 HLA-B\*7.60-84. Comp  
36 44 59.5 10 1 R83094 HLA-B\*2702 CTL modu  
37 44 59.5 10 1 R83095 HLA-B\*2702 CTL modu  
38 44 59.5 10 1 R83096 HLA-B\*2702 CTL modu  
39 44 59.5 10 1 R95425 HLA-B\*2702.75-84(D)  
40 44 59.5 10 1 R95426 HLA-B\*2702.75-84(T)  
41 44 59.5 10 1 W07513 T-cell modulating  
42 44 59.5 10 1 W47267 Immunomodulatory p  
43 44 59.5 10 1 W37269 Immunomodulatory p  
44 44 59.5 10 1 W33788 Peptide B2702.75-8  
45 44 59.5 10 1 W33787 Peptide B2702.75-8

## ALIGNMENTS

### RESULT 1

R92912  
ID R92912 standard; peptide; 15 AA.  
AC R92912;  
DT 16-MAY-1996 (first entry)  
DE HLA-B\*2702 CTL modulating peptide (B2702.70-84).  
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
KW Immunosuppressant; graft versus host disorder; transplantation; therapy;  
KW Class I MHC; HLA-B\*2702.  
OS Synthetic.  
PN W09526979-A1.  
PD 12-OCT-1995.  
PE 05-APR-1995; U04349.  
PR 05-APR-1994; US-222851.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM, Parham P;  
DR WPI; 95-358582/46.  
PT Extension of acceptance period of transplants from MHC unmatched  
PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
PT host  
PS Example 15; Page 36; 80pp; English.  
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
CC Class I major histocompatibility complex (MHC) antigens. This sequence  
CC corresponds to residues 70-84 of the alpha-1 domain of the class I MHC  
CC HLA-B\*2702. These sequences can be used to extend the period of  
CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
CC The peptides are administered to a patient in conjunction with a  
CC subtherapeutic amount of an immunosuppressant. This is administered to  
CC the patient for a limited period of time (compared to the lifetime  
CC administration for current treatments). The peptides particularly  
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
CC of the patient.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENLRALRY 15  
|||||

Db 1 KAQTDRENLRALRY 15  
|||||

### RESULT 2

W33795  
ID W33795 standard; peptide; 15 AA.  
AC W33795;  
DT 19-JUN-1998 (first entry)  
DE Peptide B2702.70-84 tested for immunomodulating activity.  
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
KW rejection.  
OS Synthetic.  
OS Homo sapiens.  
PN W09744351-A1.  
PD 27-NOV-1997.

22-MAY-1997; U08689.  
 24-MAY-1996; US-653294.  
 (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI; 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W3784-98 and W3778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 CC Sequence 15 AA;

Query Match 100.0%; Score 74; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAOTDRENLRALRY 15  
 | | | | | | | | | | | | | | | |  
 DB 1 KAOTDRENLRALRY 15

RESULT 3  
 R41205  
 ID R41205 standard; peptide; 25 AA.  
 AC R41205;  
 DT 15-MAR-1994 (first entry)  
 DE Peptide fragment of Class I HLA peptide.  
 KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
 KW parasitic disease; cytotoxic T lymphocyte; modulation.  
 OS Synthetic.  
 PN WO9317699-A.  
 PD 16-SEP-1993.  
 PF 25-FEB-1993; U01758.  
 PR 02-MAR-1992; US-844716.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger CA, Krensky AM;  
 DR WPI; 93-030134/38.  
 PT New peptide(s) based on Class I HLA antigen domains - used for  
 PT modulating cytotoxic T-lymphocyte activity towards targets  
 PS Claim 8; Page 53; 61pp; English.  
 CC The peptide (or a fragment of at least 10 amino acids, joined at at  
 CC least one terminus to a sequence other than that of wild type HLA  
 CC antigen) is used to modulate cytotoxic T-lymphocyte (CTL) activity,  
 CC either by inhibition or stimulation. It can be used for  
 CC inhibiting CTL toxicity in transplants, for inducing CTL  
 CC activity in parasitic diseases and neoplasia and in studies on viral  
 CC infection. The peptide can also be used for identifying CTLs which  
 CC bind to it and removing subsets of CTLs from a T-cell composition.  
 CC Sequence 25 AA;

Query Match 100.0%; Score 74; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. NO. 3.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAOTDRENLRALRY 15  
 | | | | | | | | | | | | | | | |  
 DB 11 KAOTDRENLRALRY 25

RESULT 4  
 R83090  
 ID R83090 standard; peptide; 25 AA.  
 AC R83090;  
 DT 16-MAY-1996 (first entry)  
 DE HLA-B2702 CTL modulating peptide (B2702.60-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 CC Class I MHC; HLA-B2702.  
 CC Synthetic.  
 OS WO9526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI; 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 13; Page 32; 80pp; English.  
 CC R83081-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
 CC Class I major histocompatibility complex (MHC) antigens. This sequence  
 CC corresponds to residues 60-84 of the alpha-1 domain of the class I MHC  
 CC HLA-B2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with  
 CC a subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 CC Sequence 25 AA;

Query Match 100.0%; Score 74; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAOTDRENLRALRY 15  
 | | | | | | | | | | | | | | | |  
 DB 11 KAOTDRENLRALRY 25

RESULT 5  
 R95416  
 ID R95416 standard; peptide; 25 AA.  
 AC R95416;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B2702.60-84.  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI; 95-194027/25.  
 PT Compns. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 9; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents

CC HLA-B2702.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC this sequence, induces calcium influx, and inhibits cytotoxic T  
 CC lymphocyte (CTL) differentiation or cytotoxicity. Candidate compounds can  
 CC be screened for their effect on the cytolytic activity of T-cells, by  
 CC combining them with the extracellular portion of p74 and determining the  
 CC amount of binding between the candidate compound and p74. Modulation of  
 CC CTL activity can be inhibited in a cellular composition containing  
 CC T-cells and antigen presenting cells (APCs), by adding to the mix the  
 CC extracellular portion of p74, in an amount sufficient to compete with p74  
 CC for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 74; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENLRALRY 15  
 |||||

Db 11 KAQTDRENLRALRY 25

RESULT 6

ID W33794 standard; peptide; 25 AA.

AC W33794; (first entry)

DE Peptide B2702.60-84 tested for immunomodulating activity.

KW Immunomodulating dimer; immunosuppressant drug; CTL activation;

KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;

KW rejection.

OS Synthetic.

PN WO9744351-A1.

PD 27-NOV-1997.

PF 22-MAY-1997; U08689.

PR 24-MAY-1996; US-653294.

PA (STRD ) UNIV LELAND STANFORD JUNIOR.

PI Beulow R, Clayberger C, Krensky AM;

DR WPI: 98-086530/08.

PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B

PT alpha-1 domain, used for preventing rejection of transplants or

PT treating autoimmune diseases

PS Example 1; Page 19; 41pp; English.

CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating

CC activity. A peptide-type compound or variant is claimed which has

CC immunomodulating activity, including the N-terminal acylated and/or

CC C-terminal amidated or esterified forms of up to 60 amino acids, where

CC the peptide-type compound comprises the formula: A-B, where A, B =

CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =

CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or

CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino

CC acid. The sequence in the brackets may optionally be absent or truncated

CC at any peptide type bond within the brackets. The compounds comprise

CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions

CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from

CC undesirably attacking cells in a host or in vitro. They can also be

CC used in combination with antigenic peptides or proteins of interest to

CC activate CTLs. They can also inhibit the proliferation of T cells in

CC response to anti-CD3. The peptide can be used for preventing rejection

CC of transplants or for treating autoimmune diseases, e.g. diabetes,

CC rheumatoid arthritis and lupus erythematosus. The products can also be

CC used for detection and diagnosis.

SQ Sequence 25 AA;

Query Match 100.0%; Score 74; DB 1; Length 25;

Best Local Similarity 100.0%; Pred. No. 3.6e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENLRALRY 15

|||||

Db 11 KAQTDRENLRALRY 25

RESULT 7

R71442

ID R71442 standard; peptide; 17 AA.

AC R71442;

DT 12-OCT-1995 (first entry)

DE Human HLA-B27-(62-85) antigen derived peptide.

KW Human HLA-B27-(62-85) antigen derived peptide; cell receptor;

KW interaction modulation; arthritis; neoplasias; lupus erythematosus.

OS Homo sapiens.

PN WO9505189-A.

PD 23-FEB-1995.

PF 12-AUG-1994; U09189.

PR 12-AUG-1993; US-105416.

PA (REGC ) UNIV CALIFORNIA.

PI Goldstein A, Goodenow RS, Olsson L;

DR WPI: 95-098577/13.

PT Regulating cell surface receptor response - by modulating

PT interaction between MHC class I antigen and the cell surface

PT receptor

PS Example 4; Page 45; 103pp; English.

CC R71439-R71443 are human major histocompatibility complex class 1

CC (MHC 1) alpha 1 domain and HLA derived peptides and fusion peptides.

CC They were used to modulate interactions between MHC 1/HLA and cell

CC surface receptors. Via competitive inhibition the peptides diminish

CC the receptors response, this feature may be useful for the treatment

CC of neoplasias, lupus erythematosus and arthritis.

SQ Sequence 17 AA;

Query Match 79.7%; Score 59; DB 1; Length 17;

Best Local Similarity 80.0%; Pred. No. 0.00012;

Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KAQTDRENLRALRY 15

|||||

Db 2 KAQTDRENLRALRY 16

RESULT 8

R41221

ID R41221 standard; peptide; 25 AA.

AC R41221;

DT 15-MAR-1994 (first entry)

DE Peptide fragment of HLA-B2705 antigen.

KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;

KW parasitic disease; cytotoxic T lymphocyte; modulation.

OS Synthetic.

PN WO9317699-A.

PD 16-SEP-1993.

PF 25-FEB-1993; U01758.

PR 02-MAR-1992; US-844716.

PA (STRD ) UNIV LELAND STANFORD JUNIOR.

PI Clayberger CA, Krensky AM;

DR WPI: 93-303134/38.

PT New peptide(s) based on Class I HLA antigen domains - used for

PT modulating cytotoxic T-lymphocyte activity towards targets

PS Example 13; Page 39; 61pp; English.

CC The peptide is used to modulate cytotoxic T-lymphocyte (CTL)

CC activity, either by inhibition or stimulation. It can be used for

CC inhibiting CTL toxicity in transplantations, for inducing CTL

CC activity in parasitic diseases and neoplasia and in studies on viral

CC infection. The peptide can also be used for identifying CTLs which

CC bind to it and removing subsets of CTLs from a T-cell composition.

CC This peptide is derived from the HLA-B2705 antigen and corresponds  
 CC to the amino acid positions 60-84 of that antigen.  
 SQ Sequence 25 AA;

Query Match 79.7%; Score 59; DB 1; Length 25;  
 Best Local Similarity 80.0%; Pred. No. 0.00019; Indels 0; Gaps 0;  
 Matches 12; Conservative 1; Mismatches 2;

OY 1 KAQTDRENRLIALRY 15  
 DB 11 KAQTDREDRLTLRY 25  
 |||||:||||

## RESULT 9

ID R83091 standard; peptide: 25 AA.  
 AC R83091:  
 DT 16-MAY-1996 (first entry)  
 DE HLA-B2702 CTL modulating peptide (B2702.60-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW class I MHC; HLA-B2702.  
 OS Synthetic.  
 PN W09526979-A1.  
 PD 12-OCT-1995. U04349.  
 PF 05-APR-1995; PF 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI: 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 13: Page 32; 80pp; English.  
 CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC corresponds to residues 60-84 of the alpha-1 domain of the class I MHC  
 CC HLA-B2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with  
 CC a subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 SQ Sequence 25 AA;

Query Match 79.7%; Score 59; DB 1; Length 25;  
 Best Local Similarity 80.0%; Pred. No. 0.00019; Indels 0; Gaps 0;  
 Matches 12; Conservative 1; Mismatches 2;

OY 1 KAQTDRENRLIALRY 15  
 DB 11 KAQTDREDRLTLRY 25  
 |||||:||||

## RESULT 10

ID R95417 standard; peptide: 25 AA.  
 AC R95417:  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B2705.60-84.  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN W09513288-A1.  
 PD 18-NOV-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.

PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Inhibit cytolytic activity and differentiation of CTLs.  
 PS Compsns. comprising lymphoid surface membrane proteins - which may  
 CC inhibit cytolytic activity and differentiation of CTLs.  
 CC Example; Page 9; 29pp; English.  
 CC R95413; and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B2705.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 79.7%; Score 59; DB 1; Length 25;  
 Best Local Similarity 80.0%; Pred. No. 0.00019;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KAQTDRENRLIALRY 15  
 DB 11 KAQTDREDRLTLRY 25  
 |||||:||||

## RESULT 11

ID P70590 standard; protein: 337 AA.  
 AC P70590:  
 DT 10-APR-1991 (first entry)  
 DE Sequence of the human histocompatibility antigen HLA B27.  
 KW Rheumatic disorder; genetic screening; diagnosis;  
 KW ankylosing spondylitis.  
 OS Homo sapiens.  
 PN DE3542024-A.  
 PD 04-JUN-1987.  
 PF 28-NOV-1985; 542024.  
 PR 28-NOV-1985; DE-542024.  
 PR 21-DEC-1985; DE-545576.  
 PA (BEHW ) BEHRINGERWERKE AG.  
 PI Riettmuller G, Meo T, Weiss E, Szots H;  
 DR WPI: 87-157893/23.  
 DR N-PSDB: N70935.  
 PT DNA coding for antigen HLA B27 - and diagnostic reagents contg.  
 PT such DNA, antigen or antibody  
 PS Disclosure; Page 5; 5pp; German.  
 CC The DNA may be used as a hybridisation probe for detecting the HLA  
 CC B27 gene, eg for assessing susceptibility to rheumatic disorders  
 CC such as ankylosis spondylitis, or may be used to transform cells  
 CC for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27  
 CC antibody in human serum, or to produce mono- or polyclonal HLA B27  
 CC antibodies for use in immunoassay.  
 SQ Sequence 337 AA;

Query Match 79.7%; Score 59; DB 1; Length 337;  
 Best Local Similarity 80.0%; Pred. No. 0.0033;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KAQTDRENRLIALRY 15



|||||:|:| |||  
70 KAQTDRDLRLTRY 84

Query Match 75.7%; Score 56; DB 1; Length 25;  
Best Local Similarity 92.3%; Pred. No. 0.00065;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## RESULT 12

ID P70155 standard; protein; 362 AA.  
AC P70155;  
DT 10-MAR-1993 (revised)  
DE 03-APR-1991 (first entry)  
DE Sequence encoded by genomic DNA encoding human histocompatibility  
DE antigen HLA-B 27.  
KW Ankylosing spondylitis; rheumatic disorder; diagnosis.  
OS Homo sapiens.  
PN EP-226069-A.  
PD 24-JUN-1987.  
PF 21-NOV-1986; 116139.  
PR 01-JAN-1985; DE-542024.  
PR 21-DEC-1985; DE-545576.  
PA (BEHW) BEHRINGERWERKE AG.  
PI Szots H, Weiss E, Dorner C, Lang M, Meo T, Riethmuller G;  
DR WPI: 87-171469/25.  
DR N-PSDB: N70225.  
DR DNA coding for human histocompatibility antigen HLA-B 27 - useful  
PT for diagnosis and antigen and antibody prodn.  
PS Disclosure; p6; 13pp; German.  
CC The DNA may be used to detect the HLA-B 27 gene (opt. mutated) in  
CC human genetic material. The HLA-B 27 may be used to detect anti-HLA-  
CC B 27 antibodies in human serum. The antibodies may be used to  
CC determine HLA-B 27 levels in human serum, eg for diagnosis of  
CC rheumatic disorders, esp. ankylosing spondylitis.  
SQ Sequence 362 AA;

Query Match 79.7%; Score 59; DB 1; Length 362;  
Best Local Similarity 80.0%; Pred. No. 0.0036;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KAQTDRDLRLTRY 15  
|||||:|:| |||  
DB 94 KAQTDRDLRLTRY 108

## RESULT 13

ID R48286 standard; peptide; 25 AA.  
AC R48286;  
DT 15-MAR-1994 (first entry)  
DE Peptide fragment of HLA-B38 antigen.  
KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
KW parasitic disease; cytotoxic T lymphocyte; modulation.  
OS Synthetic.  
PN WO9317699-A.  
PD 16-SEP-1993.  
PF 25-FEB-1993; U01758.  
PR 02-MAR-1992; US-844716.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger CA, Krensky AM;  
DR WPI: 93-303134/38.  
DR New peptide(s) based on Class I HLA antigen domains - used for  
DR modulating cytotoxic T-lymphocyte activity towards targets  
PT Example 13; Page 39; 61pp; English.  
PS The peptide is used to modulate cytotoxic T-lymphocyte (CTL)  
CC activity, either by inhibition or stimulation. It can be used for  
CC inhibiting CTL toxicity in transplantations, for inducing CTL  
CC activity in parasitic diseases and neoplasia and in studies on viral  
CC infection. The peptide can also be used for identifying CTLs which  
CC bind to it and removing subsets of CTLs from a T-cell composition.  
CC This peptide is derived from the HLA-B38 antigen and corresponds  
CC to the amino acid positions 60-84 of that antigen.  
SQ Sequence 25 AA;

QY 3 QTDRENRLRLRY 15  
|||||:|:| |||  
DB 13 QTYRENRLRLRY 25

## RESULT 14

ID R83093 standard; peptide; 25 AA.  
AC R83093;  
DT 16-MAY-1996 (first entry)  
DE HLAB38 CTL modulating peptide (B38 6084).  
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
KW class I MHC; HLAB38.  
OS Synthetic.  
PN WO9526979-A1.  
PD 12-OCT-1995.  
PF 05-APR-1995; U04349.  
PR 05-APR-1994; US-222851.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM, Parham P;  
DR WPI: 95-358582/46.  
PT Extension of acceptance period of transplants from MHC unmatched  
PT donor hosts - using Class I B7-84 MHC antigen of the recipient  
PT host  
PS Example 13; Page 32; 80pp; English.  
CC R83061-R83085, R83090-R83096 and R83097-R830913 represent fragments of  
CC class I major histocompatibility complex (MHC) antigens. This sequence  
CC corresponds to residues 60-84 of the alpha-1 domain of the class I MHC  
CC HLAB38. These sequences can be used to extend the period of acceptance  
CC by a recipient of a transplant from an MHC unmatched donor. The peptides  
CC are administered to a patient in conjunction with a subtherapeutic amount  
CC of an immunosuppressant. This is administered to the patient for a  
CC limited period of time (compared to the lifetime administration for  
CC current treatments). The peptides particularly modulate (or inhibit) the  
CC activity of the cytotoxic T lymphocytes (CTLs) of the patient.  
SQ Sequence 25 AA;

Query Match 75.7%; Score 56; DB 1; Length 25;  
Best Local Similarity 92.3%; Pred. No. 0.00065;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 QTDRENRLRLRY 15  
|||||:|:| |||  
DB 13 QTYRENRLRLRY 25

## RESULT 15

ID R95422 standard; peptide; 25 AA.  
AC R95422;  
DT 12-NOV-1996 (first entry)  
DE HLAB38.6084.  
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN WO9513288-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM;  
DR WPI: 95-194027/25.  
DR Compens. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 9; 29pp; English.

CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLAB38.6084. These sequences can be used to isolate the protein p74 from  
 CC a T-cell lysate. p74 is a T-cell surface membrane protein associated  
 CC with T-cell activation in mammalian T-cells, and is also immunologically  
 CC cross reactive with the heat shock protein Hsc70. p74 is found in a  
 CC limited number of cell types, but is particularly expressed on B and T  
 CC cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytotoxicity. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 75.7%; Score 56; DB 1; Length 25;  
 Best Local Similarity 92.3%; Pred. No. 0.00065;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 QTDRENLRALRY 15  
 || |||||  
 DB 13 QTYRENLRALRY 25

Search completed: February 8, 2000, 04:05:44  
 Job time: 9361 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 18:04:41 ; Search time 111.22 seconds  
(without alignments)  
6.362 Million cell updates/sec

Title: US-08-653-294-33

Perfect score: 74

Sequence: 1 KAQTDRENRLRY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR\_62.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	94.6	354	2 I59308	class I histocompa
2	70	94.6	354	2 I80167	class I histocompa
3	70	94.6	362	2 JH0541	class I histocompa
4	70	94.6	362	2 JH0539	class I histocompa
5	70	94.6	362	2 JH0540	class I histocompa
6	69	93.2	355	2 I80169	class I histocompa
7	69	93.2	355	2 I80171	class I histocompa
8	69	93.2	363	2 S07113	class I histocompa
9	69	93.2	363	2 S03537	class I histocompa
10	66	89.2	359	1 LHR012	MHC class I histoc
11	66	89.2	365	2 S77963	MHC class I histoc
12	66	89.2	365	2 I54416	HLA-AW24 protein -
13	66	89.2	365	2 I54493	MHC class I histoc
14	61	82.4	274	1 HLH032	MHC class I histoc
15	61	82.4	364	2 A35997	MHC class I histoc
16	60	81.1	137	2 I80174	class I histocompa
17	60	81.1	354	2 I80168	class I histocompa
18	60	81.1	362	1 LHR0B8	MHC class I histoc
19	60	81.1	362	2 I62045	gene HLA-B*1517 pr
20	60	81.1	362	2 I84490	lymphocyte antigen
21	60	81.1	362	2 I37521	HLA-B*57.2 antigen
22	60	81.1	364	2 D35997	MHC class I histoc
23	59	79.7	338	2 I56116	MHC HLA-B*27-HS - h
24	59	79.7	362	1 LHR0B2	MHC class I histoc
25	59	79.7	362	2 I37485	human lymphocyte a
26	59	79.7	362	2 I37515	MHC class I histoc
27	59	79.7	362	2 I54289	MHC HLA-B*27d - hum
28	56	75.7	273	2 I38509	MHC class I histoc
29	56	75.7	274	2 I54463	MHC HLA-B*38 chain
30	56	75.7	362	2 B30345	MHC class I histoc

31 56 75.7 362 2 A45834 MHC class I histoc  
32 56 75.7 362 2 I84486 transmembrane glyco  
33 56 75.7 362 2 A30345 MHC class I histoc  
34 56 75.7 362 2 I59633 MHC HLA-B transmem  
35 56 75.7 362 2 S24434 class I histocompa  
36 56 75.7 362 2 I37120 MHC class I histoc  
37 53 71.6 218 2 I72808 MHC class I HLA-J  
38 51 68.9 355 2 I37516 HLA-B alpha-chain  
39 51 68.9 362 2 S25415 class I histocompa  
40 51 68.9 362 2 A45850 MHC class I histoc  
41 51 68.9 362 2 I61861 MHC HLA-B\*44.2 chai  
42 51 68.9 362 2 I54442 MHC class I histoc  
43 47 63.5 365 2 JH0537 class I histocompa  
44 46 62.2 357 2 I67482 MHC class I heavy  
45 46 62.2 357 2 S11141 class I histocompa

## ALIGNMENTS

### RESULT 1

I59308

class I histocompatibility antigen - pygmy chimpanzee (fragment)

C:Species: Pan paniscus (pygmy chimpanzee, bonobo)

C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 23-Jul-1999

C:Accession: I59308

R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat

Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994

A:Title: A uniquely high level of recombination at the HLA-B locus.

A:Reference number: I59308; MUID:94286544

A:Accession: I59308

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-354 <RES>

A:Cross-references: EMBL:U05575; NID:9454767; PIDN:AAA50178.1; PID:9454768

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 94.6%; Score 70; DB 2; Length 354;

Best Local Similarity 93.3%; Pred. No. 3.5e-05;

Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENRLRY 15

Db 86 QAQTDRENRLRY 100

### RESULT 2

I80167

class I histocompatibility antigen - pygmy chimpanzee (fragment)

C:Species: Pan paniscus (pygmy chimpanzee, bonobo)

C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 23-Jul-1999

C:Accession: I80167

R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat

Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994

A:Title: A uniquely high level of recombination at the HLA-B locus.

A:Reference number: I59308; MUID:94286544

A:Accession: I80167

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-354 <RES>

A:Cross-references: EMBL:U05578; NID:9454773; PIDN:AAA50181.1; PID:9454774

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 94.6%; Score 70; DB 2; Length 354;

Best Local Similarity 93.3%; Pred. No. 3.5e-05;

Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENRLRY 15

Db 86 QAQTDRENRLRY 100

## RESULT 3

JH0541  
class I histocompatibility antigen Gogo-B0103 heavy chain precursor - lowland gorilla  
C:Species: Gorilla gorilla gorilla (lowland gorilla)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999  
C:Accession: JH0541  
R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.  
J. Exp. Med. 174, 1491-1509, 1991  
A:Title: Gorilla class I major histocompatibility complex alleles: comparison to human a  
A:Reference number: JH0534; MUID:92078860  
A:Accession: JH0541  
A:Molecule type: DNA  
A:Residues: 1-362 <LAW>  
A:Cross-references: EMBL:X60254; NID:g22869; PIDN:CAA42806.1; PID:g22870  
A:Experimental source: EBV-transformed B cell  
C:Genetics:  
A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1  
A:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-B0103 #status pre  
F:25-114/Domain: alpha-1 <AL1>  
F:115-206/Domain: alpha-2 <AL2>  
F:207-298/Domain: alpha-3 <AL3>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-362/Domain: intracellular #status predicted <INT>

Query Match 94.6%; Score 70; DB 2; Length 362;  
Best Local Similarity 93.3%; Pred. No. 3.6e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 94 QAQTDRENRLALRY 108

## RESULT 4

JH0539  
class I histocompatibility antigen Gogo-B0101 heavy chain precursor - lowland gorilla  
C:Species: Gorilla gorilla gorilla (lowland gorilla)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999  
C:Accession: JH0539  
R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.  
J. Exp. Med. 174, 1491-1509, 1991  
A:Title: Gorilla class I major histocompatibility complex alleles: comparison to human a  
A:Reference number: JH0534; MUID:92078860  
A:Accession: JH0539  
A:Molecule type: DNA  
A:Residues: 1-362 <LAW>  
A:Cross-references: EMBL:X60255; NID:g22865; PIDN:CAA42807.1; PID:g22866  
A:Experimental source: EBV-transformed B cell  
C:Genetics:  
A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1  
A:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-B0101 #status pre  
F:25-114/Domain: alpha-1 <AL1>  
F:115-206/Domain: alpha-2 <AL2>  
F:207-298/Domain: alpha-3 <AL3>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-362/Domain: intracellular #status predicted <INT>

Query Match 94.6%; Score 70; DB 2; Length 362;  
Best Local Similarity 93.3%; Pred. No. 3.6e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 94 QAQTDRENRLALRY 108

## RESULT 5

JH0540  
class I histocompatibility antigen Gogo-B0102 heavy chain precursor - lowland gorilla  
C:Species: Gorilla gorilla gorilla (lowland gorilla)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999  
C:Accession: JH0540  
R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.  
J. Exp. Med. 174, 1491-1509, 1991  
A:Title: Gorilla class I major histocompatibility complex alleles: comparison to huma  
A:Reference number: JH0534; MUID:92078860  
A:Accession: JH0540  
A:Molecule type: DNA  
A:Residues: 1-362 <LAW>  
A:Cross-references: EMBL:X60693; NID:g22867; PIDN:CAA43101.1; PID:g22868  
A:Experimental source: EBV-transformed B cell  
C:Genetics:  
A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1  
A:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-B0102 #status  
F:25-114/Domain: alpha-1 <AL1>  
F:115-206/Domain: alpha-2 <AL2>  
F:207-298/Domain: alpha-3 <AL3>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:299-362/Domain: intracellular #status predicted <INT>

Query Match 94.6%; Score 70; DB 2; Length 362;  
Best Local Similarity 93.3%; Pred. No. 3.6e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 94 QAQTDRENRLALRY 108

## RESULT 6

I80169  
class I histocompatibility antigen - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 23-Jul-1999  
C:Accession: I80169  
R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat  
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994  
A:Title: A uniquely high level of recombination at the HLA-B locus.  
A:Reference number: I59308; MUID:94286544  
A:Accession: I80169  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-355 <RES>  
A:Cross-references: EMBL:U05580; NID:g454777; PIDN:AAA50183.1; PID:g454778  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 93.2%; Score 69; DB 2; Length 355;  
Best Local Similarity 100.0%; Pred. No. 5.3e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AQTDRNRLALRY 15  
:|||||  
Db 87 AQTDRNRLALRY 100

## RESULT 7

I80171  
class I histocompatibility antigen - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 23-Jul-1999  
C:Accession: I80171  
R:McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat

Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994.  
A:Title: A uniquely high level of recombination at the HLA-B locus.

A:Reference number: 159308; MUID:94286544

A:Accession: 160171

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-355 <RES>

A:Cross-references: EMBL:U05582; NID:q454781; PID:AAA50185.1; PID:q454782

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 93.2%; Score 69; DB 2; Length 355;

Best Local Similarity 100.0%; Pred. No. 5.3e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AOTDRENLRALRY 15

|||||

Db 87 AOTDRENLRALRY 100

RESULT 8

S07113

Class I histocompatibility antigen Ch39 alpha chain precursor - chimpanzee  
N:Alternate names: MHC Ch1A chain

C:Species: Pan troglodytes (chimpanzee)

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999

C:Accession: S07113; I36957

R:Lawlor, D.A.; Ward, F.E.; Ennis, P.D.; Jackson, A.P.; Parham, P.

Nature 335, 268-271, 1988

A:Title: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees.

A:Reference number: S06424; MUID:86319000

A:Accession: S07113

A:Molecule type: mRNA

A:Residues: 1-363 <LAW>

R:Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.

J. Immunol. 142, 3937-3950, 1989

A:Title: Diversity and diversification of HLA-A,B,C alleles.

A:Reference number: I36956; MUID:89235215

A:Accession: I36957

A:Molecule type: mRNA

A:Residues: 1-363 <RES>

A:Cross-references: GB:M24045; NID:g176814; PID:AAA35424.1; PID:g176815

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: glycoprotein; membrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-363/Product: class I histocompatibility antigen Ch39 alpha chain #status predicted

F:25-114/Domain: alpha-1 #status predicted <EX1>

F:115-206/Domain: alpha-2 #status predicted <EX2>

F:220-285/Domain: immunoglobulin homology <IMM>

F:307-332/Domain: transmembrane #status predicted <TM>

F:333-363/Domain: intracellular #status predicted <INT>

F:110/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:125-188,227-283/Disulfide bonds: #status predicted

Query Match 93.2%; Score 69; DB 2; Length 363;

Best Local Similarity 100.0%; Pred. No. 5.4e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AOTDRENLRALRY 15

|||||

Db 95 AOTDRENLRALRY 108

RESULT 9

S03537

Class I histocompatibility antigen Ch1A-B1 alpha chain precursor - chimpanzee

C:Species: Pan troglodytes (chimpanzee)

C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 18-Oct-1996

C:Accession: S03537

R:Mayer, W.E.; Jonker, M.; Klein, D.; Ivanyi, P.; van Seventer, G.; Klein, J.

EMBO J. 7, 2765-2774, 1988

A:Title: Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-spec

A:Reference number: S01171; MUID:89030641

A:Accession: S03537

A:Molecule type: mRNA

A:Residues: 1-363 <MAY>

A:Cross-references: EMBL:X13115

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: transmembrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-363/Product: class I histocompatibility antigen alpha chain #status predicted <M>

F:25-298/Domain: extracellular #status predicted <EXT>

F:220-285/Domain: immunoglobulin homology <IMM>

F:299-338/Domain: transmembrane #status predicted <TM>

F:339-363/Domain: intracellular #status predicted <INT>

Query Match 93.2%; Score 69; DB 2; Length 363;

Best Local Similarity 100.0%; Pred. No. 5.4e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AOTDRENLRALRY 15

|||||

Db 95 AOTDRENLRALRY 108

RESULT 10

HLH012

MHC class I histocompatibility antigen HLA alpha chain precursor (clone PHLA 12.4) -

C:Species: Homo sapiens (man)

C:Date: 05-Apr-1983 #sequence\_revision 05-Apr-1983 #text\_change 22-Jun-1999

C:Accession: A02189

R:Malissen, M.; Malissen, B.; Jordan, B.R.

Proc. Natl. Acad. Sci. U.S.A. 79, 893-897, 1982

A:Title: Exon/Intron organization and complete nucleotide sequence of an HLA gene.

A:Reference number: A02189; MUID:82151002

A:Accession: A02189

A:Molecule type: DNA

A:Residues: 1-359 <MAL>

A:Cross-references: GB:J00191; GB:V00526; NID:g187600; PID:AAA36218.1; PID:g386873

C:Comment: The seven exons correspond approximately to the domain structure of this c

C:Genetics:

A:Map position: 6p21.3

A:Introns: 22/1; 112/1; 204/1; 296/1; 335/1; 346/1

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterodimer; transmembrane protein; transplant

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-359/Product: class I histocompatibility antigen HLA alpha chain #status predicted

F:22-304/Domain: extracellular #status predicted <EXT>

F:22-111/Domain: alpha-1 <EX1>

F:112-203/Domain: immunoglobulin homology <IMM>

F:217-282/Domain: transmembrane #status predicted <TM>

F:305-329/Domain: intracellular #status predicted <INT>

F:107/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:224-280/Disulfide bonds: #status predicted

Query Match 89.2%; Score 66; DB 1; Length 359;

Best Local Similarity 86.7%; Pred. No. 0.00019;

Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KAOTDRENLRALRY 15

|||||

Db 91 KAOTDRENLRALRY 105

RESULT 11

S77963

MHC class I histocompatibility antigen HLA-A alpha chain precursor - human (isolate A

C:Species: Homo sapiens (man)

A:Variety: isolate A2402

C:Date: 13-Jan-1995 #sequence\_revision 15-Aug-1997 #text\_change 23-Jul-1999

C:Accession: S77963

R:Belich, M.P.; Madrigal, J.A.; Hildebrand, W.H.; Zemmour, J.; Williams, R.C.; Luz, R

Nature 357, 326-329, 1992  
 A:Title: Unusual HLA-B alleles in two tribes of Brazilian Indians.  
 A:Reference number: 137120; MUID:92289955  
 A:Accession: S77963  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-365 <BEL>  
 A:Cross-references: EMBL:M64740; NID:g403142; PIDN:AAA59600.1; PID:g187614  
 A:Experimental source: cell line KRC 032; isolate A\*2402  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1991  
 C:Genetics:  
 A:Gene: GDB:HLA-A  
 A:Cross-references: GDB:119310; OMIM:142800  
 A:Map position: 6p21.3-6p21.3  
 C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
 C:Keywords: transmembrane protein  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-365/Product: MHC class I histocompatibility antigen HLA-A alpha chain #status predicted  
 F:220-285/Domain: immunoglobulin homology <IMM>

Query Match 89.2%; Score 66; DB 2; Length 365;  
 Best Local Similarity 92.9%; Pred. No. 0.00019;  
 Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENLRALRY 15  
 :|||||  
 Db 95 SQTDRNLRLALRY 108

RESULT 12  
 I54416  
 HLA-AW24 protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
 C:Accession: I54416  
 R:N'Guyen, C.; Sodoyez, R.; Trucy, J.; Strachan, T.; Jordan, B.R.  
 Immunogenetics 21, 479-489, 1985  
 A:Title: The HLA-AW24 gene: sequence, surroundings and comparison with the HLA-A2 and HL  
 A:Reference number: I54416; MUID:85206128  
 A:Accession: I54416  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-365 <RES>  
 A:Cross-references: GB:M15497; NID:g187644; PIDN:AAA59611.1; PID:g386877  
 C:Genetics: 1-365 <RES>  
 A:Introns: 25/1; 115/1; 207/1; 338/1; 349/1; 365/1  
 C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 89.2%; Score 66; DB 2; Length 365;  
 Best Local Similarity 92.9%; Pred. No. 0.00019;  
 Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENLRALRY 15  
 :|||||  
 Db 95 SQTDRNLRLALRY 108

RESULT 13  
 I54493  
 MHC class I histocompatibility antigen HLA-A alpha chain precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
 C:Accession: I54493; I55659  
 R:Little, A.M.; Madrigal, J.A.; Parham, P.  
 Immunogenetics 35, 41-45, 1992  
 A:Title: Molecular definition of an elusive third HLA-A9 molecule: HLA-A9.3.  
 A:Reference number: I54493; MUID:92104637  
 A:Accession: I54493  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-365 <RES>

A:Cross-references: GB:M64741; NID:g187615; PIDN:AAA59601.1; PID:g187616  
 R:Kragel, M.S.  
 J. Exp. Med. 163, 1173-1190, 1986  
 A:Title: Secretion of HLA-A and -B antigens via an alternative RNA splicing pathway.  
 A:Reference number: I55659; MUID:86198522  
 A:Accession: I55659  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: mRNA  
 A:Residues: 271-365 <RE2>  
 A:Cross-references: GB:M12377; NID:g187607; PIDN:AAA17889.1; PID:g187608  
 C:Genetics:  
 A:Gene: GDB:HLA-A  
 A:Cross-references: GDB:119310; OMIM:142800  
 A:Map position: 6p21.3-6p21.3  
 A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1  
 C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
 C:Keywords: glycoprotein

Query Match 89.2%; Score 66; DB 2; Length 365;  
 Best Local Similarity 92.9%; Pred. No. 0.00019;  
 Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENLRALRY 15  
 :|||||  
 Db 95 SQTDRNLRLALRY 108

RESULT 14  
 HLHU32  
 MHC class I histocompatibility antigen HLA-A32 alpha chain - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 05-Sep-1997  
 C:Accession: A26088  
 R:Wan, A.M.; Ennis, P.; Parham, P.; Holmes, N.  
 J. Immunol. 137, 3671-3674, 1986  
 A:Title: The primary structure of HLA-A32 suggests a region involved in formation of  
 A:Reference number: A26088; MUID:87058961  
 A:Accession: A26088  
 A:Molecule type: protein  
 A:Residues: 1-274 <WAN>  
 C:Genetics:  
 A:Gene: GDB:HLA-A  
 A:Cross-references: GDB:119310; OMIM:142800  
 A:Map position: 6p21.3-6p21.3  
 C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
 C:Keywords: glycoprotein; heterodimer; transmembrane protein; transplantation antigen  
 F:196-261/Domain: immunoglobulin homology <IMM>  
 F:86/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 82.4%; Score 61; DB 1; Length 274;  
 Best Local Similarity 85.7%; Pred. No. 0.0011;  
 Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENLRALRY 15  
 :|||||  
 Db 71 SQTDRNLRLALRY 84

RESULT 15  
 A35997  
 MHC class I histocompatibility antigen HLA-A25 alpha chain precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 16-Nov-1990 #sequence\_revision 16-Nov-1990 #text\_change 16-Feb-1997  
 C:Accession: A35997  
 R:Ennis, P.D.; Zemmour, J.; Salter, R.D.; Parham, P.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 2833-2837, 1990  
 A:Title: Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: freque  
 A:Reference number: A35997; MUID:90207291  
 A:Accession: A35997  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA

A:Residues: 1-364 <ENN>  
A:Cross-references: GB:M32321  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:219-284/Domain: immunoglobulin homology <IMM>

Query Match 82.4%; Score 61; DB 2; Length 364;  
Best Local Similarity 85.7%; Pred. No. 0.0015;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENRLALRY 15  
:||||:|||||  
Db 95 SQTDRSLALRY 108

Search completed: February 7, 2000, 18:04:41  
Job time: 22207 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:26:01 ; Search time 68.63 seconds  
(without alignments)  
6.527 Million cell updates/sec

Title: US-08-653-294-33

Perfect score: 74

Sequence: 1 KAQTDRENLRALRY 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_38:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	74	100.0	362	1 IB15_HUMAN	P10317 homo sapien
2	70	94.6	362	1 IB01_GORGO	P30379 gorilla gor
3	70	94.6	362	1 IB02_GORGO	P30380 gorilla gor
4	70	94.6	362	1 IB03_GORGO	P30381 gorilla gor
5	69	93.2	359	1 IB01_PANTR	P13750 pan troglod
6	66	89.2	362	1 HLAB_HUMAN	P01893 homo sapien
7	66	89.2	365	1 IA23_HUMAN	P30447 homo sapien
8	66	89.2	365	1 IA24_HUMAN	P05534 homo sapien
9	61	82.4	365	1 IA25_HUMAN	P18462 homo sapien
10	61	82.4	365	1 IA32_HUMAN	P10314 homo sapien
11	60	81.1	362	1 IB60_HUMAN	P18465 homo sapien
12	60	81.1	362	1 IB61_HUMAN	P30497 homo sapien
13	60	81.1	362	1 IB62_HUMAN	P10319 homo sapien
14	59	79.7	338	1 IB20_HUMAN	P30467 homo sapien
15	59	79.7	361	1 IB14_HUMAN	P03989 homo sapien
16	59	79.7	362	1 IB16_HUMAN	P19373 homo sapien
17	59	79.7	362	1 IB18_HUMAN	P10318 homo sapien
18	59	79.7	362	1 IB19_HUMAN	Q08136 homo sapien
19	56	75.7	362	1 IB47_HUMAN	P30487 homo sapien
20	56	75.7	362	1 IB49_HUMAN	P18464 homo sapien
21	56	75.7	362	1 IB52_HUMAN	P30489 homo sapien
22	56	75.7	362	1 IB53_HUMAN	P30490 homo sapien
23	56	75.7	362	1 IB54_HUMAN	P30491 homo sapien
24	51	68.9	362	1 IB05_HUMAN	P30461 homo sapien
25	51	68.9	362	1 IB41_HUMAN	P30481 homo sapien
26	51	68.9	362	1 IB42_HUMAN	P30482 homo sapien
27	47	63.5	365	1 IA04_GORGO	P30378 gorilla gor
28	46	62.2	359	1 IB40_HUMAN	P10320 homo sapien
29	43	58.1	361	1 HA1A_RABIT	P01894 oryctolagus
30	43	58.1	361	1 HA1B_RABIT	P06140 oryctolagus
31	41	55.4	362	1 IB02_HUMAN	P01889 homo sapien
32	41	55.4	362	1 IB29_HUMAN	P18463 homo sapien
33	41	55.4	362	1 IB39_HUMAN	P30480 homo sapien
34	41	55.4	362	1 IB45_HUMAN	P30485 homo sapien

## RESULT 1

ID	IB15_HUMAN	STANDARD;	PRT;	362 AA.	
AC	P10317:				
DT	01-MAR-1989	(Rel. 10, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	01-APR-1993	(Rel. 25, Last annotation update)			
DE	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2702 ALPHA CHAIN				
DE	PRECURSOR (B-27K) (B27.2).				
GN	HLA-B OR HLAB.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eukaryota; Primates; Catarrhini; Hominidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 86220133.				
RA	SEEMANN G.H.A.; REIN R.S.; BROWN C.S.; PLOEGH H.L.;				
RT	"Gene conversion-like mechanisms may generate polymorphism in human				
RT	class I genes.";				
RL	EMBO J. 5:547-552(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	PARRAM P.; ARNETT K.L.; ADAMS E.J.;				
RL	Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 86-107 AND 171-181.				
RX	MEDLINE; 86042671.				
RA	VEGA M.A.; EZQUERRA A.; ROJO S.; APARICIO P.; BRAGADO R.;				
RA	LOPEZ DE CASTRO J.A.;				
RT	"Structural analysis of an HLA-B27 functional variant: identification				
RT	of residues that contribute to the specificity of recognition by				
RT	cytolytic T lymphocytes.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 82:7394-7398(1985).				
CC	- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO				
CC	THE IMMUNE SYSTEM.				
CC	- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-				
CC	MICROGLOBULIN).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; X03664; CAA27301.1; .				
DR	EMBL; X03667; CAA27301.1; JOINED.				
DR	EMBL; L38504; AAA69724.1; .				
DR	PIR; B25092; HLHUBK.				
DR	HSP; P03989; 1HSA.				
DR	MIM; 142830; .				
DR	PROSITE; P500290; IG_MHC; 1.				
DR	PFAM; PF00047; ig; 1.				
DR	PFAM; PF00129; MHC_I; 1.				
KW	MHC I; Transmembrane; Glycoprotein; Signal.				
FT	SIGNAL				

FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 B-27 B\*2702 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 308 CONNECTING PEPTIDE.  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40397 MW; 9798F0BB CRC32;

Query Match 100.0%; Score 74; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KAOTDRENLRALRY 15  
 :|||||  
 DB 94 KAOTDRENLRALRY 108

RESULT 2  
 ID 1B01\_GORGO STANDARD; PRT; 362 AA.  
 AC P30379; (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0101 ALPHA CHAIN PRECURSOR.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92078860.  
 RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;  
 RT "Gorilla class I major histocompatibility complex alleles: comparison  
 to human and chimpanzee class I.";  
 RL J. Exp. Med. 174:1491-1509(1991).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X60255; CAA42807.1; -  
 DR PIR; JH0539; JH0539.  
 DR HSP; P03989; LHSA.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 362 CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 GOGO-B0101 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 308 CONNECTING PEPTIDE.  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 FT CARBOHYD 110 110 BY SIMILARITY.

SQ SEQUENCE 362 AA; 40170 MW; 2E33E2B8 CRC32;  
 Query Match 94.6%; Score 70; DB 1; Length 362;  
 Best Local Similarity 93.3%; Pred. No. 1.4e-05;  
 Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KAOTDRENLRALRY 15  
 :|||||  
 DB 94 KAOTDRENLRALRY 108

RESULT 3  
 ID 1B02\_GORGO STANDARD; PRT; 362 AA.  
 AC P30380; (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0102 ALPHA CHAIN PRECURSOR.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92078860.  
 RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;  
 RT "Gorilla class I major histocompatibility complex alleles: comparison  
 to human and chimpanzee class I.";  
 RL J. Exp. Med. 174:1491-1509(1991).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X60693; CAA43101.1; -  
 DR PIR; JH0540; JH0540.  
 DR HSP; P03989; LHSA.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 362 CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 GOGO-B0102 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 308 CONNECTING PEPTIDE.  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40204 MW; 3CF119AD CRC32;

Query Match 94.6%; Score 70; DB 1; Length 362;  
 Best Local Similarity 93.3%; Pred. No. 1.4e-05;  
 Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KAOTDRENLRALRY 15  
 :|||||  
 DB 94 KAOTDRENLRALRY 108

```

RESULT 4
1B03_GORGO STANDARD; PRT; 362 AA.
AC P30381;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0103 ALPHA CHAIN PRECURSOR.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92078860.
RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT "Gorilla class I major histocompatibility complex alleles: comparison
to human and chimpanzee class I.";
RL J. Exp. Med. 174:1491-1509(1991).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60254; CAA42806.1; -
DR PIR; JH0541; JH0541.
DR HSP; P03989; IHSA.
DR PFAM; PF00047; IG_MHC; 1.
DR PFAM; PF00129; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT CLASS I HISTOCOMPATIBILITY ANTIGEN,
GOGO-B0103 ALPHA CHAIN.
FT DOMAIN 25 114
EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206
EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298
EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308
CONNECTING PEPTIDE.
FT TRANSMEM 309 332
CYTOPLASMIC TAIL.
FT DOMAIN 333 362
BY SIMILARITY.
FT DISULFID 125 188
BY SIMILARITY.
FT DISULFID 227 283
BY SIMILARITY.
FT CARBOHYD 110 110
BY SIMILARITY.
SQ SEQUENCE 362 AA; 40248 MW; FEA6A941 CRC32;

Query Match 94.6%; Score 70; DB 1; Length 362;
Best Local Similarity 93.3%; Pred. No. 1.4e-05;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENLRALRY 15
Db 94 QAQTDRENLRALRY 108
:|||||
RESULT 5
1B01_PANTR STANDARD; PRT; 359 AA.
AC P13750;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN PRECURSOR
(FRAGMENT).
OS Pan troglodytes (Chimpanzee).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89030641.
RA KLEIN J.;
RA KLEIN J.;
RT "Nucleotide sequences of chimpanzee MHC class I alleles: evidence for
trans-species mode of evolution.";
RL EMBO J. 7:2765-2774(1988).
RN [2]
RP REVISIONS.
RA MAYER W.;
RA Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13115; CAA31507.1; -
DR PIR; S03537; S03537.
DR HSP; P03989; IHSA.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; IG; 1.
DR PFAM; PF00129; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 359
FT CLASS I HISTOCOMPATIBILITY ANTIGEN,
B-1 ALPHA CHAIN.
FT DOMAIN 21 110
EXTRACELLULAR ALPHA-1.
FT DOMAIN 111 202
EXTRACELLULAR ALPHA-2.
FT DOMAIN 203 294
EXTRACELLULAR ALPHA-3.
FT DOMAIN 295 305
CONNECTING PEPTIDE.
FT TRANSMEM 306 329
CYTOPLASMIC TAIL.
FT DOMAIN 330 359
BY SIMILARITY.
FT DISULFID 121 184
BY SIMILARITY.
FT DISULFID 223 279
BY SIMILARITY.
FT CARBOHYD 106 106
BY SIMILARITY.
SQ SEQUENCE 359 AA; 40173 MW; 5395FFC9 CRC32;

Query Match 93.2%; Score 69; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTDRNLRALRY 15
Db 91 AQTDRNLRALRY 104
|||||
RESULT 6
HLAH_HUMAN STANDARD; PRT; 362 AA.
AC P01893;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN H PRECURSOR
(HLA-A*) (HLA-12.4).
GN HLA-H OR HLAH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE: 82151002.
RA MALISSEN M., MALISSEN B., JORDAN B.R.:
RT "Exon/Intron organization and complete nucleotide sequence of an HLA
RL gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:893-897(1982).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM. COULD BE THE PRODUCT OF A PSEUDOGENE.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J00191; AAA36218.1; ALT_INIT.
DR PIR: A02189; HLHUL2.
DR HSP: P03989; IHSA.
DR MIM: 142800; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_1; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT ALPHA CHAIN H.
FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308 CONNECTING PEPTIDE.
FT TRANSMEM 309 332
FT DOMAIN 333 362
FT CARBOHYD 110 110 CYTOPLASMIC TAIL.
FT DISULFID 227 283 BY SIMILARITY.
SQ SEQUENCE 362 AA; 40850 MW; 56610F63 CRC32;

Query Match 89.2%; Score 66; DB 1; Length 362;
Best Local Similarity 86.7%; Pred. No. 7.8e-05;
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAOTDRENLRIALRY 15
DB 94 QAOTERENLRIALRY 108
:|||||:|||||

RESULT 7
ID 1A23_HUMAN STANDARD; PRT; 365 AA.
AC P30447;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-23(A-9) ALPHA CHAIN
DE PRECURSOR.
GN HLA-A OR HLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN SEQUENCE FROM N.A. (A*2301).
RX MEDLINE: 92104637.
RA LITTLE A.-M., MADRIGAL J.A., PARHAM P.;
RT "Molecular definition of an elusive third HLA-A9 molecule: HLA-A9.3.";
RL Immunogenetics 35:41-45(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).

```

```

CC -!- POLYMORPHISM: THE ONLY ALLELE OF A-23 KNOWN IS A*2301 WHICH IS
CC SHOWN HERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M64742; AAA03662.1; -.
DR HSP: P01892; IAQD.
DR MIM: 142800; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_1; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 365 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT A-23(A-9) ALPHA CHAIN.
FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308 CONNECTING PEPTIDE.
FT TRANSMEM 309 332
FT DOMAIN 333 365
FT CARBOHYD 110 110 CYTOPLASMIC TAIL.
FT DISULFID 227 283 BY SIMILARITY.
SQ SEQUENCE 365 AA; 40732 MW; B1C21094 CRC32;

Query Match 89.2%; Score 66; DB 1; Length 365;
Best Local Similarity 92.9%; Pred. No. 7.9e-05;
Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENLRIALRY 15
DB 95 SOTDRENLRIALRY 108
:|||||:|||||

RESULT 8
ID 1A24_HUMAN STANDARD; PRT; 365 AA.
AC P05534; P30448; P30449; Q95355;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-24(A-9) ALPHA CHAIN
DE PRECURSOR (AW-24).
GN HLA-A OR HLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN SEQUENCE FROM N.A. (A*2401).
RX MEDLINE: 85206128.
RA N'GUYEN C., SODOYER R., TRUCY J., STRACHAN T., JORDAN B.R.;
RT "The HLA-AW24 gene: sequence, surroundings and comparison with the
RL HLA-A2 and HLA-A3 genes.";
RL Immunogenetics 21:479-489(1985).
[2]
RN REVISIONS (A*2401).
RA JORDAN B.R.;
RL Submitted (XXX-1988) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A. (A*2402/A*2403).
RX MEDLINE: 92104637.
RA LITTLE A.-M., MADRIGAL J.A., PARHAM P.;
RT "Molecular definition of an elusive third HLA-A9 molecule: HLA-A9.3.";
RL Immunogenetics 35:41-45(1992).
[4]

```

```

RP SEQUENCE FROM N.A. (A*2402).
RX MEDLINE: 92269955.
RA BELICH M.P., MADRICAL J.A., HILDEBRAND W.H., ZEMMOUR J.,
RA WILLIAMS R.C., LUZ R., PETZL-ERLER M.L., PARHAM P.;
RT "Unusual HLA-B alleles in two tribes of Brazilian Indians.";
RL Nature 357:326-329(1992).
RN [5]
RP SEQUENCE FROM N.A. (A*2402).
RA LAFORET M., TONGIO M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 26-206 FROM N.A.
RA GAO X., MCCLUSKEY J.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A. (A*2408).
RC TISSUE-BLOOD;
RA KASHIWASE K., TOKUNAGA K., ISHIKAWA Y., OIU L., FURUYA M.,
RA SAWANAKA K., AKAZA T., TADOKORO K., JUJI T.;
RT "A new A9 sequence HLA-A*9HH from Japanese.";
RL MHC 3:9-14(1996).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -1- POLYMORPHISM: THE FOLLOWING ALLELES OF A*24 ARE KNOWN: A*2401,
CC A*2401, A*2403 AND A*2408 (A9HH). THE SEQUENCE SHOWN IS THAT OF
CC A*2401.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15497; AAA59611.1; -
CC DR EMBL: M64740; AAA59600.1; -
CC DR EMBL: M64741; AAA59601.1; -
CC DR EMBL: U19733; AAB60651.1; -
CC DR EMBL: U18987; AAB60651.1; JOINED.
CC DR EMBL: Z72422; CAA96532.1; -
CC DR EMBL: D83516; BAA11936.1; -
CC DR HSP: P01892; IAQD.
CC MW: 142800; -
CC PROSITE: PS00290; IG_MHC; 1.
CC PFAM: PF00047; Ig; 1.
CC DR PFAM: PF00129; MHC_I; 1.
CC MW: 142800; -
CC MHC I; Transmembrane; Glycoprotein; Signal; Polymorphism.
CC -----
CC SIGNAL 1 24
CC CHAIN 25 365 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC A-24(A-9) ALPHA CHAIN.
CC DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
CC FT TRANSMEM 309 332 CONNECTING PEPTIDE.
CC FT DOMAIN 333 365 CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110 BY SIMILARITY.
CC FT DISULFID 125 188 BY SIMILARITY.
CC FT DISULFID 227 283 BY SIMILARITY.
CC FT VARIANT 5 5 G->A (IN A*2402, A*2403 AND A*2408).
CC FT VARIANT 27 27 /FTID=VAR_004354.
CC FT VARIANT 86 86 H->Q (IN A*2408).
CC FT VARIANT 89 89 E->G (IN A*2408).
CC FT VARIANT 180 180 /FTID=VAR_004356.
CC FT VARIANT 180 180 G->R (IN A*2408).
CC FT VARIANT 180 180 /FTID=VAR_004357.
CC FT VARIANT 180 180 Q->W (IN REF. 6).
CC FT VARIANT 190 191 /FTID=VAR_004358.
CC FT VARIANT 191 191 DG->EW (IN A*2403).
CC -----
FT FT /FTID=VAR_004359.
FT A->T (IN A*2402, A*2403, A*2408 AND
FT REF. 6).
FT /FTID=VAR_004360.
SQ SEQUENCE 365 AA; 40644 MW; DE23D06E CRC32;
Query Match 89.2%; Score 66; DB 1; Length 365;
Best Local Similarity 92.9%; Pred. No. 7.9e-05;
Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 AQTDRNLRIALRY 15
DB 95 SQTDRNLRIALRY 108
RESULT 9
ID 1A25_HUMAN STANDARD; PRT; 365 AA.
AC P18462; Q95362;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 13-DEC-1999 (Rel. 39, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-25(A-10) ALPHA CHAIN
DE PRECURSOR.
GN HLA-A OR HLA*.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RN SEQUENCE FROM N.A. (A*2501).
RX MEDLINE: 90207291.
RA ENNIS P.D., ZEMMOUR J., SALTER R.D., PARHAM P.;
RT "Rapid cloning of HLA-A,B CDNA by using the polymerase chain
RT reaction: frequency and nature of errors produced in amplification.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2833-2837(1990).
RN [2]
RP SEQUENCE FROM N.A. (A*2502).
RC TISSUE-HEMATOPOIETIC;
RX MEDLINE: 97045042.
RA KRAUSA P., YOUNG D.M., GOTCH F.;
RT "Identification of a new HLA-A allele (A*2502) by PCR-SSP.";
RL Immunogenetics 45:84-85(1996).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -1- POLYMORPHISM: THE FOLLOWING ALLELES OF A-25 ARE KNOWN: A*2501
CC AND A*2502. THE SEQUENCE SHOWN IS THAT OF A*2501.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M32321; AAA36234.1; -
CC DR EMBL: X97802; CAA66389.1; -
CC PIR: A35997; A35997.
CC HSP: P01891; 2HLA.
CC MW: 142800; -
CC PROSITE: PS00290; IG_MHC; 1.
CC PFAM: PF00047; Ig; 1.
CC DR PFAM: PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal; Polymorphism.
CC -----
CC SIGNAL 1 24
CC CHAIN 25 365 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC A-25(A-10) ALPHA CHAIN.
CC DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.

```

```

FT DOMAIN 299 308 CONNECTING PEPTIDE.
FT TRANSMEM 309 332
FT DOMAIN 333 365 CYTOPLASMIC TAIL.
FT CARBOHYD 110 110 BY SIMILARITY.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
FT VARIANT 94 94 H -> Q (IN A*2502).
FT SEQUENCE 365 AA; 41218 MW; 131A8F20 CRC32;

Query Match 82.4%; Score 61; DB 1; Length 365;
Best Local Similarity 85.7%; Pred. No. 0.00066;
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTDRNLRLRY 15
DB 95 SQTDRSLRLRY 108

RESULT 10
1A32_HUMAN
ID 1A32_HUMAN STANDARD; PRT; 365 AA.
AC P10314; Q29838;
DT 01-NOV-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-32(AW-19) ALPHA CHAIN
DE PRECURSOR.
GN HLA-A OR HLA-A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (A*3201).
RA DOMENA J.D.;
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 25-298 FROM N.A. (A*3201).
RX MEDLINE; 87058961.
RA WAN A.M., ENNIS P., PARHAM P., HOLMES N.;
RT "The primary structure of HLA-A32 suggests a region involved in
RT formation of the Bw4/Bw6 epitopes.";
RL J. Immunol. 137:3671-3674(1986).
RN [3]
RP SEQUENCE FROM N.A. (A*3202).
RC TISSUE-BLOOD;
RX MEDLINE; 97045038.
RA ZINO E., SEVERINI G.M., MAZZI B., BORDIGNON C., BENAZZI E.,
RA FLEISCHHAUER K.;
RT "Sequencing of a new HLA-A*32 subtype (A*3202).";
RL Immunogenetics 45:76-77(1996).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -!- POLYMORPHISM: THE FOLLOWING ALLELES OF A-32 ARE KNOWN: A*3201 AND
CC AND A*3202. THE SEQUENCE SHOWN IS THAT OF A*3201.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U03907; AAA03605.1; -
CC EMBL; X97120; CA65786.1; -
CC PIR; A26088; HLH032.
CC HSP; P01892; 1HHH.
CC MIM; 142800; -
CC PROSITE; PS00290; IG_MHC; 1.

```

```

DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal; Polymorphism.
FT SIGNAL 1 24
FT CHAIN 25 365
FT DOMAIN 25 114 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 115 206 A-32(AW-19) ALPHA CHAIN.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-1.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-2.
FT TRANSMEM 309 332 EXTRACELLULAR ALPHA-3.
FT DOMAIN 333 365 CONNECTING PEPTIDE.
FT CARBOHYD 110 110 CYTOPLASMIC TAIL.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
FT VARIANT 175 175 R -> H (IN A*3202).
FT VARIANT 180 180 /FTID=VAR_004373.
FT SEQUENCE 365 AA; 41048 MW; 71FA71A9 CRC32;

Query Match 82.4%; Score 61; DB 1; Length 365;
Best Local Similarity 85.7%; Pred. No. 0.00066;
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTDRNLRLRY 15
DB 95 SQTDRSLRLRY 108

RESULT 11
1B60_HUMAN
ID 1B60_HUMAN STANDARD; PRT; 362 AA.
AC P18465;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B*57(B-17) B*5701 ALPHA
DE CHAIN PRECURSOR (B*57.1).
GN HLA-B OR HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90207291.
RA ENNIS P.D., ZEMMOUR J., SALTER R.D., PARHAM P.;
RT "Rapid cloning of HLA-A,B CDNA by using the polymerase chain
RT reaction: frequency and nature of errors produced in amplification.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2833-2837(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91067476.
RA ISAMAT M., GIRDLESTONE J., MILSTEIN C.;
RT "Nucleotide sequence of an HLA-B*57 gene.";
RL Nucleic Acids Res. 18:6702-6702(1990).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M32318; AAA36231.1; -
CC EMBL; X55711; CAA39244.1; -
CC PIR; S12622; S12622.

```

DR PIR: D35997; D35997.  
 DR HSP: P30491; 1A1M.  
 DR MW: 142830; -.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 DR MHC I; Transmembrane; Glycoprotein; Signal.  
 KW SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 B-57(B-17) B\*5701 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 308 CONNECTING PEPTIDE.  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40224 MW; D91DF8DD CRC32;

Query Match 81.1%; Score 60; DB 1; Length 362;  
 Best Local Similarity 92.9%; Pred. No. 0.001;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AQTDRNLRIALRY 15  
 DB 95 AQTRENLRIALRY 108  
 III IIIIIIIIIII

RESULT 12  
 1B61\_HUMAN  
 ID 1B61\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30497;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-57(B-17) B\*5702 ALPHA CHAIN  
 DE PRECURSOR (BW57.2).  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93056508.  
 RA MADRIGAL J.A., BELICH M.P., HILDEBRAND W.H., BENJAMIN R.J.,  
 RA LITTLE A.M., ZEMMOUR J., ENNIS P.D., WARD F.E., PETZL-ERLER M.L.,  
 RA MARTEL R.W., DU TOIT E.D., PARHAM P.;  
 FT "Distinctive HLA-A,B antigens of black populations formed by  
 RT interallelic conversion.";  
 RL J. Immunol. 149:3411-3415(1992).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X61707; CAA43876.1; -.  
 DR PIR: S16774; S16774.  
 DR HSP: P30491; 1A1M.  
 DR MW: 142830; -.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.

KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 B-57(B-17) B\*5702 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 308 CONNECTING PEPTIDE.  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40342 MW; 628C2156 CRC32;

Query Match 81.1%; Score 60; DB 1; Length 362;  
 Best Local Similarity 92.9%; Pred. No. 0.001;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AQTDRNLRIALRY 15  
 DB 95 AQTRENLRIALRY 108  
 III IIIIIIIIIII

RESULT 13  
 1B62\_HUMAN  
 ID 1B62\_HUMAN STANDARD; PRT; 362 AA.  
 AC P10319;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-58(B-17) B\*5801 ALPHA  
 DE CHAIN PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 86008247.  
 RA WAYS J.P., COPPIN H.L., PARHAM P.;  
 RT "The complete primary structure of HLA-Bw58.";  
 RL J. Biol. Chem. 260:11924-11933(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BLOOD;  
 RA INQUE T., OGAWA A.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M11799; AAS59628.1; -.  
 DR EMBL: AB008102; BAA22916.1; -.  
 DR PIR: A23895; HLHUB8.  
 DR HSP: P30491; 1A1M.  
 DR MW: 142830; -.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; Ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,

FT DOMAIN 25 114 BW-58(B-17) B\*5801 ALPHA CHAIN.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40337 MW; 3E5E7534 CRC32;

Query Match 81.1%; Score 60; DB 1; Length 362;  
 Best Local Similarity 92.9%; Pred. No. 0.001; 1; Indels 0; Gaps 0;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AQTDRNLRLALRY 15  
 ||| ||||| |||||  
 DB 95 AQTRENLRLALRY 108

RESULT 14  
 1B20\_HUMAN STANDARD; PRT; 338 AA.  
 AC P30467;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B\*2707 ALPHA CHAIN  
 DE (B27-HS).  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91268545.  
 RA CHOO Y.S., FAN L.A., HANSEN J.A.;  
 RT "A novel HLA-B27 allele maps B27 allospecificity to the region around  
 RT position 70 in the alpha 1 domain.";  
 RL J. Immunol. 147:174-180(1991).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: M62852; AAA59647.1; --  
 DR HSP: P03989; LISA.  
 DR MIM: 142830; --  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 90 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 91 182 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 183 274 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 275 284 CONNECTING PEPTIDE.  
 FT TRANSMEM 285 308  
 FT DOMAIN 309 338 CYTOPLASMIC TAIL.  
 FT CARBOHYD 86 86 BY SIMILARITY.  
 FT DISULFID 101 164 BY SIMILARITY.  
 FT DISULFID 203 259 BY SIMILARITY.  
 SQ SEQUENCE 338 AA; 37804 MW; 33FB8134 CRC32;

Query Match 79.7%; Score 59; DB 1; Length 338;  
 Best Local Similarity 80.0%; Pred. No. 0.0014; 2; Indels 0; Gaps 0;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KAQTDRENRLALRY 15  
 ||||| ||||| |||||  
 DB 70 KAQTDRENRLALRY 84

RESULT 15  
 1B14\_HUMAN STANDARD; PRT; 361 AA.  
 AC P03989;  
 DT 23-OCT-1986 (Rel. 02, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 ALPHA CHAIN PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86138405.  
 RA WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.;  
 RT "Organization, sequence and expression of the HLA-B27 gene: a  
 RT molecular approach to analyze HLA and disease associations.";  
 RL Immunobiology 170:367-380(1985).  
 RN [2]  
 RP SEQUENCE OF 25-361 FROM N.A.  
 RX MEDLINE: 86149317.  
 RA SZOETS H., RIETHMUELLER G., WEISS E., MEO T.;  
 RT "Complete sequence of HLA-B27 CDNA identified through the  
 RT characterization of structural markers unique to the HLA-A, -B, and  
 RT -C allelic series.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1428-1432(1986).  
 RN [3]  
 RP SEQUENCE OF 25-295.  
 RX MEDLINE: 85246361.  
 RA EZQUERRA A., BRAGADO R., VEGA M.A., STROMINGER J.L., WOODY J.,  
 RA LOPEZ DE CASTRO J.A.;  
 RT "Primary structure of papain-solubilized human histocompatibility  
 RT antigen HLA-B27.";  
 RL Biochemistry 24:1733-1741(1985).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 25-300.  
 RX MEDLINE: 92405152.  
 RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;  
 RT "The three-dimensional structure of HLA-B27 at 2.1-A resolution  
 RT suggests a general mechanism for tight peptide binding to MHC.";  
 RL Cell 70:1035-1048(1992).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX MEDLINE: 92018187.  
 RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;  
 RT "The structure of HLA-B27 reveals nonamer self-peptides bound in an  
 RT extended conformation.";  
 RL Nature 353:321-325(1991).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -!- DISEASE: THIS PROTEIN CORRELATES WITH THE DEVELOPMENT OF  
 CC ANKYLOSING SPONDYLITIS.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



CC -----  
DR EMBL: X03945; CAA27578.1; ALT\_TERM.  
DR PIR: A25128; HLHUB2.  
DR PIR: S07441; S07441.  
DR PDB: 1HSA; 15-OCT-92.  
DR MIM: 142830; -.  
DR PROSITE: PS00290; IG\_MHC; 1.  
DR PFAM: PF00047; Ig; 1.  
DR PFAM: PF00129; MHC\_I; 1.  
KW MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 25 361  
FT DOMAIN 25 114  
FT DOMAIN 115 206  
FT DOMAIN 207 298  
FT DOMAIN 299 308  
FT TRANSMEM 309 332  
FT DOMAIN 333 361  
FT CARBOHYD 110 110  
FT DISULFID 125 188  
FT DISULFID 227 283  
FT CONFLICT 206 206  
FT CONFLICT 266 266  
FT STRAND 27 38  
FT TURN 39 41  
FT STRAND 42 52  
FT TURN 53 54  
FT STRAND 55 61  
FT TURN 62 63  
FT STRAND 70 71  
FT HELIX 74 76  
FT TURN 77 78  
FT HELIX 81 108  
FT TURN 109 110  
FT TURN 113 114  
FT STRAND 118 127  
FT TURN 129 130  
FT STRAND 133 142  
FT TURN 143 144  
FT STRAND 145 150  
FT TURN 152 153  
FT STRAND 157 159  
FT HELIX 162 173  
FT TURN 174 175  
FT HELIX 176 185  
FT TURN 186 186  
FT HELIX 187 198  
FT TURN 199 199  
FT HELIX 200 203  
FT TURN 204 204  
FT STRAND 207 207  
FT STRAND 210 217  
FT STRAND 222 233  
FT STRAND 238 243  
FT TURN 244 245  
FT STRAND 246 247  
FT HELIX 249 251  
FT STRAND 253 254  
FT STRAND 258 259  
FT STRAND 265 274  
FT TURN 275 276  
FT HELIX 278 280  
FT STRAND 281 286  
FT TURN 288 289  
FT STRAND 294 296  
SQ SEQUENCE 361 AA; 802130D5 CRC32;

QY 1 KAQTDRENRLIALRY 15  
||| ||| |||  
Db 94 KAQTDREDLRLILRY 108

Search completed: February 8, 2000, 01:36:01  
Job time: 1561 sec

Query Match 79.7%; Score 59; DB 1; Length 361;  
Best Local Similarity 80.0%; Pred. NO. 0.0015;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 19:16:17 ; Search time 176.54 Seconds  
(without alignments)  
5.891 Million cell updates/sec

Title: US-08-653-294-33

Perfect score: 74

Sequence: 1 KAOTDRENLRALRY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL12:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	133	7 019189	019189 homo sapien
2	70	94.6	181	7 019354	019354 gorilla gor
3	70	94.6	354	7 095524	095524 pan paniscu
4	70	94.6	354	7 095527	095527 pan paniscu
5	69	93.2	354	7 095412	095412 hylobates 1
6	69	93.2	355	7 095529	095529 pan troglod
7	69	93.2	363	7 030988	030988 pan troglod
8	67	90.5	180	7 019607	019607 homo sapien
9	67	90.5	180	7 019608	019608 homo sapien
10	67	90.5	180	7 019609	019609 homo sapien
11	67	90.5	180	7 019610	019610 homo sapien
12	67	90.5	181	7 029724	029724 homo sapien
13	67	90.5	365	7 029689	029689 homo sapien
14	66	89.2	90	7 046697	046697 gorilla gor
15	66	89.2	138	7 078209	078209 homo sapien
16	66	89.2	180	7 019611	019611 homo sapien
17	66	89.2	180	7 019612	019612 homo sapien
18	66	89.2	180	7 019613	019613 homo sapien
19	66	89.2	181	7 062892	062892 homo sapien
20	66	89.2	181	7 062920	062920 homo sapien

21	66	89.2	181	7 062923	062923 homo sapien
22	66	89.2	181	7 029910	029910 homo sapien
23	66	89.2	181	7 019521	019521 homo sapien
24	66	89.2	181	7 029909	029909 homo sapien
25	66	89.2	181	7 029841	029841 homo sapien
26	66	89.2	181	7 077937	077937 homo sapien
27	66	89.2	181	7 029908	029908 homo sapien
28	66	89.2	181	7 098012	098012 homo sapien
29	66	89.2	182	7 079484	079484 homo sapien
30	66	89.2	200	7 078168	078168 homo sapien
31	66	89.2	345	7 015506	015506 homo sapien
32	63	85.1	90	7 046693	046693 pan troglod
33	63	85.1	90	7 046694	046694 pan troglod
34	63	85.1	90	7 046695	046695 pan troglod
35	63	85.1	90	7 046696	046696 pan paniscu
36	63	85.1	359	7 030595	030595 macaca mula
37	60	81.1	137	7 095533	095533 pan troglod
38	60	81.1	181	7 062917	062917 homo sapien
39	60	81.1	181	7 019631	019631 homo sapien
40	60	81.1	181	7 019769	019769 homo sapien
41	60	81.1	181	7 028679	028679 homo sapien
42	60	81.1	181	7 029733	029733 homo sapien
43	60	81.1	181	7 097997	097997 homo sapien
44	60	81.1	354	7 095528	095528 pan troglod
45	60	81.1	359	7 095353	095353 homo sapien

#### ALIGNMENTS

RESULT 1  
019189 PRELIMINARY; PRT; 133 AA.  
ID 019189  
AC 019189;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
DE MHC CLASS I HISTOCOMPATIBILITY ANTIGEN-B (HLA-B-27KSH) (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEUKOCYTE;  
RA PETERSDORF E.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U18659; AAB60357.1; -.  
DR MIM; 142830; -.  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC I.  
FT NON\_TER 1  
FT NON\_TER 133  
SQ SEQUENCE 133 AA; 15491 MW; 3A3BC802 CRC32;

Query Match 100.0%; Score 74; DB 7; Length 133;  
Best local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAOTDRENLRALRY 15  
|||||  
DB 22 KAOTDRENLRALRY 36

RESULT 2  
019354 PRELIMINARY; PRT; 181 AA.  
ID 019354  
AC 019354;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
DE MHC CLASS I A ANTIGEN (FRAGMENT).

OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Gorilla.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94286544.  
 RA CEREB N., KONG Y., LEE S., MAYE P., YANG S.Y.;  
 RT "Nucleotide sequences of MHC class I introns 1, 2, and 3 in humans and  
 RL Tissue Antigens 47:498-511(1996).  
 DR EMBL; U37323; AAB49915.1; -;  
 DR HSSP; P10318; 1ROG.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 181 181  
 SQ SEQUENCE 181 AA; 20981 MW; FFB51D78 CRC32;

Query Match 94.6%; Score 70; DB 7; Length 181;  
 Best Local Similarity 93.3%; Pred. No. 4.7e-05;  
 Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KAOTDRENLRALRY 15  
 Db 69 QAOTDRENLRALRY 83

RESULT 3  
 Q95524  
 ID Q95524 PRELIMINARY; PRT; 354 AA.  
 AC Q95524;  
 DT 01-FEB-1997 (TRENBLrel. 02, Created)  
 DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).  
 GN HLA-B.  
 OS Pan paniscus (Pygmy chimpanzee) (Bonobo).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-LOREL;  
 RX MEDLINE; 94286544.  
 RA MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,  
 RA BONTROP R.E., WATKINS D.I.;  
 RT "A uniquely high level of recombination at the HLA-B locus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5893-5897(1994).  
 DR EMBL; U05575; AAA50178.1; -;  
 DR HSSP; P30491; 1A1M.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 39227 MW; 5D75939D CRC32;  
 SQ SEQUENCE 354 AA; 39227 MW; 5D75939D CRC32;

Query Match 94.6%; Score 70; DB 7; Length 354;  
 Best Local Similarity 93.3%; Pred. No. 9.6e-05;  
 Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KAOTDRENLRALRY 15  
 Db 86 QAOTDRENLRALRY 100

RESULT 4  
 Q95527  
 ID Q95527 PRELIMINARY; PRT; 354 AA.  
 AC Q95527;  
 DT 01-FEB-1997 (TRENBLrel. 02, Created)  
 DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)

DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).  
 GN HLA-B.  
 OS Pan paniscus (Pygmy chimpanzee) (Bonobo).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-BOSONDJO;  
 RX MEDLINE; 94286544.  
 RA MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,  
 RA BONTROP R.E., WATKINS D.I.;  
 RT "A uniquely high level of recombination at the HLA-B locus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5893-5897(1994).  
 DR EMBL; U05578; AAA50181.1; -;  
 DR HSSP; P30491; 1A1M.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 354 AA; 39401 MW; 6E38CFFE CRC32;  
 SQ SEQUENCE 354 AA; 39401 MW; 6E38CFFE CRC32;

Query Match 94.6%; Score 70; DB 7; Length 354;  
 Best Local Similarity 93.3%; Pred. No. 9.6e-05;  
 Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KAOTDRENLRALRY 15  
 Db 86 QAOTDRENLRALRY 100

RESULT 5  
 Q95412  
 ID Q95412 PRELIMINARY; PRT; 354 AA.  
 AC Q95412;  
 DT 01-FEB-1997 (TRENBLrel. 02, Created)  
 DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).  
 GN HLA-B.  
 OS Hylobates lar (Common gibbon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC MEDLINE; 92218857.  
 RX CHEN Z.W., MCADAM S.N., HUGHES A.L., DOGON A.L., LETVIN N.L.,  
 RA WATKINS D.I.;  
 RT "Molecular cloning of orangutan and gibbon MHC class I cDNA. The HLA-A  
 RT and -B loci diverged over 30 million years ago."  
 RL J. Immunol. 148:2547-2554(1992).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC CHEN Z.W., MCADAM S.N., HUGHES A.L., WATKINS D.I.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U50091; AAB08074.1; -;  
 DR HSSP; P30491; 1A1M.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 354 AA; 39748 MW; 43EDEC6F CRC32;

Query Match 93.2%; Score 69; DB 7; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AOTDRENLRALRY 15  
 Db 87 QAOTDRENLRALRY 100

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTDRNLRLRY 15  
 DB 95 AQTDRNLRLRY 108

RESULT 8

ID 019607 PRELIMINARY; PRT; 180 AA.  
 AC 019607;  
 DT 01-JAN-1998 (TREMELrel. 05, Created)  
 DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE MHC CLASS I HLA-A (FRAGMENT).  
 GN HLA-A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF030920; AAB87056.1; -;  
 DR EMBL; AF030919; AAB87056.1; JOINED.  
 DR HSSP; P01891; 1TMC.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 180  
 SQ SEQUENCE 180 AA; 20811 MW; CECC3537 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 180;  
 Best Local Similarity 86.7%; Pred. No. 0.00016;  
 Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAQTDRENRLRY 15  
 DB 69 QSQTDRENRLRY 83

RESULT 9

ID 019608 PRELIMINARY; PRT; 180 AA.  
 AC 019608;  
 DT 01-JAN-1998 (TREMELrel. 05, Created)  
 DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE MHC CLASS I HLA-A (FRAGMENT).  
 GN HLA-A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF030922; AAB87057.1; -;  
 DR EMBL; AF030921; AAB87057.1; JOINED.  
 DR HSSP; P01891; 1TMC.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 180  
 SQ SEQUENCE 180 AA; 20811 MW; CECC3537 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 180;  
 Best Local Similarity 86.7%; Pred. No. 0.00016;  
 Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
 Q95529 PRELIMINARY; PRT; 355 AA.  
 AC Q95529;  
 DT 01-FEB-1997 (TREMELrel. 02, Created)  
 DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).  
 GN HLA-B.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-TEPPIE;  
 RX MEDLINE; 94286544.  
 RA MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,  
 RA BONTROP R.E., WATKINS D.I.;  
 RT "A uniquely high level of recombination at the HLA-B locus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5893-5897(1994).  
 DR EMBL; U05580; AAA50183.1; -;  
 DR HSSP; P30491; 1A1M.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 355  
 SQ SEQUENCE 355 AA; 39772 MW; 8AD2B984 CRC32;

Query Match 93.2%; Score 69; DB 7; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTDRNLRLRY 15  
 DB 87 AQTDRNLRLRY 100

RESULT 7

ID Q30988 PRELIMINARY; PRT; 363 AA.  
 AC Q30988;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE CHIMPANZEE MHC CLASS I CHLA CHAIN.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RA ENNIS P.D.;  
 RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 23-363 FROM N.A.  
 RX MEDLINE; 89235215.  
 RA FARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
 RT "Diversity and diversification of HLA-A,B,C alleles.";  
 RL J. Immunol. 142:3937-3950(1989).  
 DR EMBL; M24045; AAA35424.1; -;  
 DR HSSP; P30491; 1A1M.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 SQ SEQUENCE 363 AA; 40540 MW; 133979D1 CRC32;

Query Match 93.2%; Score 69; DB 7; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 0.00015;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 69 QSQTDRENRLALRY 83

RESULT 10  
O19609 PRELIMINARY; PRT; 180 AA.  
AC O19609;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE MHC CLASS I HLA-A (FRAGMENT).  
GN HLA-A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF030924; AAB87425.1; -.  
DR HSSP; AF030923; AAB87425.1; JOINED.  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 180 180  
SQ SEQUENCE 180 AA; 20811 MW; CECC3537 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 180;  
Best Local Similarity 86.7%; Pred. No. 0.00016;  
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 69 QSQTDRENRLALRY 83

RESULT 11  
O19610 PRELIMINARY; PRT; 180 AA.  
AC O19610;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE MHC CLASS I HLA-A (FRAGMENT).  
GN HLA-A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF030926; AAB87058.1; -.  
DR EMBL; AF030925; AAB87058.1; JOINED.  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 180 180  
SQ SEQUENCE 180 AA; 20811 MW; CECC3537 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 180;  
Best Local Similarity 86.7%; Pred. No. 0.00016;  
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||

Db 69 QSQTDRENRLALRY 83  
:|||||

RESULT 12  
Q29724 PRELIMINARY; PRT; 181 AA.  
AC Q29724;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-A (FRAGMENT).  
GN HLA-A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHOPEK M., NIELSEN J., ZHANG H.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U36914; AAA80238.1; -.  
DR HSSP; P01891; ITMC.  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 181 181  
SQ SEQUENCE 181 AA; 20912 MW; D475FCD7 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 181;  
Best Local Similarity 86.7%; Pred. No. 0.00016;  
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 69 QSQTDRENRLALRY 83

RESULT 13  
Q29689 PRELIMINARY; PRT; 365 AA.  
AC Q29689;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-A2407.  
GN HLA-A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SELVAKUMAR A.;  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25971; AAA73518.1; -.  
DR HSSP; P01892; IHHH.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PFAM; PF00047; Ig; 1.  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
SQ SEQUENCE 365 AA; 40679 MW; A23C2F19 CRC32;

Query Match 90.5%; Score 67; DB 7; Length 365;  
Best Local Similarity 86.7%; Pred. No. 0.00034;  
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAQTDRENRLALRY 15  
:|||||  
Db 94 QSQTDRENRLALRY 108

RESULT 14

O46697  
 ID O46697 PRELIMINARY; PRT; 90 AA.  
 AC O46697;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-H ORTHOLOG (FRAGMENT).  
 GN HLA-H.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SHAMBA;  
 RA GRIMSLEY C., MATHER K.A., OBER C.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022172; AAC99794.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 FT NON\_TER 90 90  
 SQ SEQUENCE 90 AA; 10689 MW; 5E5F2495 CRC32;

Query Match 89.2%; Score 66; DB 7; Length 90;  
 Best Local Similarity 86.7%; Pred. No. 0.00011;  
 Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAOTDRENLRALRY 15  
 :||:|||||||  
 Db 70 QAOTERENLRALRY 84

## RESULT 15

O78209  
 ID O78209 PRELIMINARY; PRT; 138 AA.  
 AC O78209;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE HUMAN LEUCOCYTE ANTIGEN PRECURSOR (FRAGMENT).  
 GN HLA-A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 9800772.  
 RA LAFORET M., FROELICH N., PARISSADIS A., BAUSINGER H., PFEIFFER B.,  
 RA TONGIO M.M.;  
 RT "An intronic mutation responsible for a low level of expression of an  
 RT HLA-A\*24 allele."  
 RL Tissue Antigens 50:340-346(1997).  
 DR EMBL; Z72423; CAA96533.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW Signal; MHC.  
 FT SIGNAL 1 24  
 FT NON\_TER 138 138  
 SQ SEQUENCE 138 AA; 15610 MW; B8417FA0 CRC32;

Query Match 89.2%; Score 66; DB 7; Length 138;  
 Best Local Similarity 92.9%; Pred. No. 0.00018;  
 Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AOTDRENLRALRY 15  
 :|||||||  
 Db 95 SOTDRENLRALRY 108

**THIS PAGE BLANK (USPTO)**



OM of: US-08-653-294-33 to: GenEmbl.\* out\_format : pfs

Date: Feb 8, 2000 10:25 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: p2n.model -DEV-xlp  
-Q-cnrl\_1/USPTO.spool/US08653294/runat\_04022000\_160701\_15779/app\_query.fasta.2  
-DB-GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-33

Query length: 15

Database: GenEmbl.\*

Database sequences: 821193

Database length: -1518192014

Search time (sec): 10176.920000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_pr2:HSU18659	74.00	255.18	9.5e-07	399	U18659 Human MHC class I HLA-B
gb_pr2:HSU18659	74.00	255.22	3.0e-06	1089	L38504 Homo sapiens (clones 18
gb_pr2:HSU18659	74.00	255.06	3.5e-06	1241	X03664 Human class I MHC gene
gb_pr2:HSU18659	70.00	243.82	1.5e-05	792	U37323 Gorilla gorilla MHC clas
gb_pr2:HSU18659	70.00	241.18	2.1e-05	1065	U05575 Pan paniscus class I h
gb_pr2:HSU18659	70.00	241.18	2.1e-05	1065	U05578 Pan paniscus class I h
gb_pr2:HSU18659	70.00	239.61	2.5e-05	1270	X60255 G.gorilla MhcGogo-B0101
gb_pr2:HSU18659	70.00	239.61	2.5e-05	1270	X50693 G.gorilla MhcGogo-B0102
gb_pr2:HSU18659	70.00	239.61	2.5e-05	1270	X60254 G.gorilla MhcGogo-B0103
gb_pr2:HSU18659	69.00	237.37	3.4e-05	1085	U50091 Hylobates lar MHC class
gb_pr2:HSU18659	69.00	237.35	3.4e-05	1088	U05580 Pan troglodytes class I
gb_pr2:HSU18659	69.00	237.35	3.4e-05	1088	U05582 Pan troglodytes class I
gb_pr2:HSU18659	69.00	237.17	3.5e-05	1089	AF115460 Pan troglodytes isolate
gb_pr2:HSU18659	69.00	234.27	5.0e-05	1507	X13115 Chimpanzee mRNA for cla
gb_pr2:HSU18659	69.00	234.10	5.1e-05	1537	M24045 Chimpanzee MHC class I
gb_pr2:HSU18659	67.00	241.99	1.9e-05	270	L43530 Homo sapiens (clone K920
gb_pr2:HSU18659	67.00	241.99	1.9e-05	270	AF030923 Homo sapiens MHC class
gb_pr2:HSU18659	67.00	241.99	1.9e-05	270	AF030921 Homo sapiens MHC class
gb_pr2:HSU18659	67.00	241.99	1.9e-05	270	AF030919 Homo sapiens MHC class
gb_pr2:HSU18659	67.00	241.99	1.9e-05	270	AF030925 Homo sapiens MHC class
gb_pr2:HSU18659	67.00	241.99	1.9e-05	270	AF140723 Homo sapiens MHC class
gb_pr2:HSU18659	67.00	235.71	4.2e-05	546	U35914 Human MHC class I antigen
gb_pr2:HSU18659	67.00	229.48	9.9e-05	1098	U25971 Human MHC class I antigen
gb_pr2:HSU18659	66.00	238.86	2.8e-05	250	Y08692 H.sapiens HLA-B gene, ex
gb_pr2:HSU18659	66.00	238.86	2.8e-05	250	Y08693 H.sapiens HLA-B gene, ex
gb_pr2:HSU18659	66.00	238.69	2.8e-05	255	AF022159 Homo sapiens isolate C
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	I14590 Sequence 67 from patent
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	I14591 Sequence 68 from patent
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	I14592 Sequence 69 from patent
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	I14593 Sequence 70 from patent
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	X82161 H.sapiens HLA-A gene, al
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	AJ011699 Homo sapiens HLA-A gen
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	AJ239035 Homo sapiens HLA-A gen
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	U37110 Human HLA-A24 gene, alle
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	U37114 Human HLA-A24 gene, alle
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	U37112 Human HLA-A24 gene, alle
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	I14987 Human MHC class I antigen
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	L43528 Homo sapiens (clone K620
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	AF030931 Homo sapiens MHC class
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	AF067436 Homo sapiens MHC class
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	AF030911 Homo sapiens MHC class
gb_pr2:HSU18659	66.00	238.18	3.0e-05	270	AF030927 Homo sapiens MHC class

gb\_pr3:HS639HLA1 + 66.00 238.18 3.0e-05 270 ! AF030929 Homo sapiens MHC cl  
gb\_pr3:HS639HLA1 + 66.00 238.18 3.0e-05 270 ! AF030915 Homo sapiens MHC cl  
gb\_pr4:HS639HLA1 + 66.00 238.18 3.0e-05 270 ! U59699 Human MHC class I (HL

seq\_name: gb\_pr2:HSU18659

seq\_documentation\_block:

LOCUS HSU18659 399 bp mRNA PRI 30-APR-1995  
DEFINITION Human MHC class I HLA-B mRNA (HLA-B-27KSH allele) exons 2 and 3,  
partial cds.  
ACCESSION U18659  
VERSION U18659.1 GI:790214  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 399)  
AUTHORS Petersdorf,E.  
TITLE Molecular Diversity of HLA-B  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 399)  
AUTHORS Petersdorf,E.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-1994) Effie Petersdorf, Human Immunogenetics,  
Fred Hutchinson Cancer Research Center, 1124 Columbia, Seattle, WA  
98104, USA

FEATURES

Location/Qualifiers

1..399  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.3"  
/cell\_type="leukocyte"  
/tissue\_type="blood"  
1..399  
/gene="HLA-B"  
<1..127  
/gene="HLA-B"  
/number=2  
<1..>399  
/gene="HLA-B"  
/notes="HLA-B-27KSH allele"  
/codon\_start=1  
/protein\_id="PAB60357.1"  
/db\_xref="GI:790215"  
/translation="APWIEQGEYWDRETQICKAKAOTDRENLRALRYYNQSEAGS  
HTLQNNYGDVGPDRGLRGYHQDAYDKGYIALNEDLSSWTAAADTAQAQITQKWEAA  
RYAEQLRAYLEGCEVWLRRLRYLENGKETLQR"  
128..>399  
/gene="HLA-B"  
/number=3  
372  
/gene="HLA-B"  
/notes="This submission reports an HLA-B-2702 variant  
(HLA-B-27KSH) that differs from the published B-2702  
sequence. GenBank Accession Number X03664"  
/phenotype="HLA-B-27"  
/replace="t"

BASE COUNT 87 a 119 c 145 g 48 t  
ORIGIN

alignment\_scores:  
Quality: 74.00 Length: 15  
Ratio: 4.933 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-33 x HSU18659

Align seg 1/1 to: HSU18659 from: 1 to: 399

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
64 AAGGCACAGACTGACCGAGAGACCTGCGGATCGCGCTCGCTAC 108

seq\_name: gb\_pr2:HUMHLABC

seq\_documentation\_block:  
LOCUS HUMHLABC 1089 bp mRNA PRI 11-JUL-1995  
DEFINITION Homo sapiens (clones 18.1, 18.2, 19.2) MHC class I HLA-B\*2702 mRNA,  
complete cds.  
ACCESSION L38504  
VERSION L38504.1 GI:896270  
KEYWORDS cell surface antigen; cell surface glycoprotein; class I gene;  
integral membrane protein; major histocompatibility complex.  
SOURCE Homo sapiens (clone: B18.1) cDNA to mRNA; Homo sapiens (clone:  
18.2) cDNA to mRNA; and Homo sapiens (clone: 19.2) cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1089)  
AUTHORS Parham,P., Arnett,K.L. and Adams,E.J.  
TITLE On the nucleotide sequences of B\*2702 and B\*2705  
JOURNAL Unpublished (1995)  
FEATURES  
Location/Qualifiers  
1..1089  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="B18.1"  
/cell\_line="NV"  
/cell\_type="EBV-transformed B-cell"  
/haplotype="HLA A2/3; B7; Cw2,7"  
/clone="18.2"  
/clone="19.2"  
/map="6p21.3"  
1..1089  
/gene="HLA-B\*2702"  
1..1089  
/gene="HLA-B\*2702"  
/product="major histocompatibility complex"  
/protein\_id="AAA69724.1"  
/db\_xref="GI:896271"  
/translation="MRVTAPRTLLLLWGAVALTETWAGSHSMRYFHTSVSRGGRGEP  
RFTVGYDDTLFVRFDSDAASPREAPRWIEQGEPEYWDRETQICKAKAQDRENL  
RIALRYNQSEAGSHTLQNMVCGDVGDRLLRGYHODAYDGKDYIALNEDLSWTAA  
DTAQITQRKEARVAEQRLRAYLEGCEVWLRRLYENGKETLQADPPKTHVTHPI  
SDHEATLKWALGFPAEITLTWRDGEDQDTLVELVETRPAGDRTQKWAAYVVPSSG  
EEQRYTCHVOHEGLPKPLTLRWEPSSQSTVPIGIVAGLAVLVVIGAVVAVMCRR  
KSSGKGGSYSQAACSDSAQSDVSLTA"  
<1..73

exon  
exon  
exon  
exon  
exon  
exon  
exon

BASE COUNT 217 a 332 c 368 g 172 t

ORIGIN

alignment\_scores:  
Quality: 74.00 Length: 15  
Ratio: 4.933 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-33 x HUMHLABC ..  
Align seg 1/1 to: HUMHLABC from: 1 to: 1089

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
280 AAGGCACAGACTGACCGAGAGACCTGCGGATCGCGCTCGCTAC 324

seq\_name: gb\_pr1:HSLLAKI

seq\_documentation\_block:  
LOCUS HSLLAKI 1241 bp DNA PRI 28-JAN-1997  
DEFINITION Human class I MHC gene HLA-B\*27K exons 1-3 (BRUG cell line).  
ACCESSION X03664  
VERSION X03664.1 GI:32236  
KEYWORDS class I antigen; histocompatibility antigen; major  
histocompatibility complex; signal peptide.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1241)  
AUTHORS Seemann,G.H., Rein,R.S., Brown,C.S. and Ploegh,H.L.  
TITLE Gene conversion-like mechanisms may generate polymorphism in human  
class I genes  
JOURNAL EMBO J. 5 (3), 547-552 (1986)  
MEDLINE 86220133  
COMMENT 380 bp of intron III missing. Sequence continued in X03667.  
FEATURES  
Location/Qualifiers  
1..1241  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="BRUG"  
join(<79..151,280..549,793..1068,X03667.1:6..281,  
X03667.1:373..489,X03667.1:930..962,X03667.1:1089..1116,  
X03667.1:1299..>1639)  
/label="HLA\_mRNA"  
<79..151  
/gene="HLA-B\*27K"  
/label="ex1"  
join(79..151,280..549,793..1068,X03667.1:6..281,  
X03667.1:373..489,X03667.1:930..962,X03667.1:1089..1112).  
/gene="HLA-B\*27K"  
join(79..151,280..549,793..1068,X03667.1:6..281,  
X03667.1:373..489,X03667.1:930..962,X03667.1:1089..1112)  
/gene="HLA-B\*27K"  
/codon\_start=1  
/label="HLA\_CDS"  
/db\_xref="GI:871296"  
/db\_xref="SWISS-PROT:P10317"  
/translation="MRVTAPRTLLLLWGAVALTETWAGSHSMRYFHTSVSRGGRGEP  
RFTVGYDDTLFVRFDSDAASPREAPRWIEQGEPEYWDRETQICKAKAQDRENL  
RIALRYNQSEAGSHTLQNMVCGDVGDRLLRGYHODAYDGKDYIALNEDLSWTAA  
DTAQITQRKEARVAEQRLRAYLEGCEVWLRRLYENGKETLQADPPKTHVTHPI  
SDHEATLKWALGFPAEITLTWRDGEDQDTLVELVETRPAGDRTQKWAAYVVPSSG  
EEQRYTCHVOHEGLPKPLTLRWEPSSQSTVPIGIVAGLAVLVVIGAVVAVMCRR  
KSSGKGGSYSQAACSDSAQSDVSLTA"  
280..549  
/gene="HLA-B\*27K"  
/label="ex2"  
793..1068  
/gene="HLA-B\*27K"  
/label="ex3"  
BASE COUNT 213 a 399 c 447 g 182 t

ORIGIN

alignment\_scores:  
Quality: 74.00 Length: 15  
Ratio: 4.933 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

## alignment\_block:

US-08-653-294-33 x HSHLAK1

Align seg 1/1 to: HSHLAK1 from: 1 to: 1241

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
486 AAGGCACAGACTGACCGAGAGAACCTGCGGATCGCGCTCGCTAC 530

seq\_name: gb\_pr2:GGU37323

## seq\_documentation\_block:

LOCUS GGU37323 792 bp DNA PRI 21-MAR-1997  
DEFINITION Gorilla gorilla MHC class I B antigen gene, partial cds.  
ACCESSION U37323  
VERSION U37323.1 GI:1654122  
KEYWORDS  
SOURCE gorilla.  
ORGANISM Gorilla gorilla  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
REFERENCE 1 (bases 1 to 792)  
AUTHORS Cereb.N., Kong.Y., Lee.S., Maye.P. and Yang.S.Y.  
TITLE Nucleotide sequences of MHC class I introns 1, 2, and 3 in humans and intron 2 in nonhuman primates  
JOURNAL Tissue Antigens 47 (6), 498-511 (1996)  
MEDLINE 96408732  
REFERENCE 2 (bases 1 to 792)  
AUTHORS Yang,S.Y. and Cereb.N.  
TITLE Direct Submission  
JOURNAL Submitted (29-SEP-1995) Soo Yang, Immunology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, Box 41, New York, NY 10021, USA

## FEATURES

Location/Qualifiers

1..792  
/organism="Gorilla gorilla"  
/db\_xref="taxon:9593"  
/cell\_line="ROK B cell line"  
/clone="ROK1/2"  
/chromosome="5"  
join(<1..270,517..>792)  
/codon\_start=3  
/product="MHC class I A antigen"  
/protein\_id="AAB49915.1"  
/db\_xref="GI:1654123"  
/translation="SHSMRYEDTAVSRPGCEPRFIVGVYDDTQVRFDSDAASPRM  
EPRAWIEGPEYWDRETQTSKAQADRENLRALRYYNQSEAGSHTIQWMTGCDM  
GPDGRLLRGYSQAYDGKDYIALNEDLSWTAAADTAQAQITQKWEAAAEAEQLRAYLE  
GTCVEWLRRLRYLENGRETQLQA"

## exon

1..270

/number=2

271..516

/number=2

517..792

/number=3

## BASE COUNT

141 a 263 c 284 g 104 t

## alignment\_scores:

Quality: 70.00 Length: 15  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 93.333

## alignment\_block:

US-08-653-294-33 x GGU37323

Align seg 1/1 to: GGU37323 from: 1 to: 792

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
207 CAGGCACAGACTGACCGAGAGAACCTGCGGATCGCGCTCGCTAC 251

seq\_name: gb\_pr1:PPU05575

## seq\_documentation\_block:

LOCUS PPU05575 1065 bp mRNA PRI 08-OCT-1994  
DEFINITION Pan paniscus class I histocompatibility antigen Papa-B (Papa-B-03 allele) mRNA, partial cds.  
ACCESSION U05575  
VERSION U05575.1 GI:454767  
KEYWORDS  
SOURCE pygmy chimpanzee.  
ORGANISM Pan paniscus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE 1 (bases 1 to 1065)  
AUTHORS McAdam,S.N., Boyson,J.E., Liu,X., Garber,T.L., Hughes,A.L., Bontrop,R.E. and Watkins,D.I.  
TITLE A uniquely high level of recombination at the HLA-B locus  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (13), 5893-5897 (1994)  
MEDLINE 94286544  
REFERENCE 2 (bases 1 to 1065)  
AUTHORS Boyson,J.E.  
TITLE Direct Submission  
JOURNAL Submitted (25-JAN-1994) Jonathan E. Boyson, Watkins Unit, Wisconsin Regional Primate Research Center, 1220 Capitol Court, Madison, WI 53715, USA

## FEATURES

Location/Qualifiers

1..1065  
/organism="Pan paniscus"  
/isolate="Lorel"  
/db\_xref="taxon:9597"  
1..1065  
/gene="HLA-B"  
<1..1065  
/gene="HLA-B"  
/note="HLA-B-03 allele"  
/codon\_start=1  
/product="class I histocompatibility antigen"  
/protein\_id="AAA50178.1"  
/db\_xref="GI:454768"  
/translation="VLLLSAALATETWAGSHSMRYEDTAVSRPGAGEPRFIVGVY  
DDTQVRFDSDAASPREPRAPWMEQGEPEYWDRETQTSKAQADRENLRALRYYN  
QSEAGSHTIQWMTGCDMGPGRLLRGYSQAYDGKDYIALNEDLSWTAAADTAQAQITQ  
KWEAAAEAEQLRAYLENGRETQLQAQITQKWEAAAEAEQLRAYLENGRETQLQA  
CWALGFYPAEITLTWORDGDTQDTLQVETRPAGDRTFKWAAVVPVSGOEQRVYTC  
VQHGLEPILTLRWEPSSQSTPIPIGVIVAGLVAVVTGAVVAAMVCRKSSGGKGG  
SYSQAASDSDAQSDVSLTA"

## BASE COUNT

208 a 328 c 361 g 168 t

## alignment\_scores:

Quality: 70.00 Length: 15  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 93.333

## alignment\_block:

US-08-653-294-33 x PPU05575

Align seg 1/1 to: PPU05575 from: 1 to: 1065

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
256 CAGGCACAGACTGACCGAGAGAACCTGCGGATCGCGCTCGCTAC 300

seq\_name: gb\_pr1:PPU05578

## seq\_documentation\_block:

LOCUS PPU05578 1065 bp mRNA PRI 08-OCT-1994  
DEFINITION Pan paniscus class I histocompatibility antigen Papa-B (Papa-B-02 allele) mRNA, partial cds.  
ACCESSION U05578  
VERSION U05578.1 GI:454773  
KEYWORDS

SOURCE pygmy chimpanzee.  
ORGANISM Pan paniscus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1 (bases 1 to 1065)  
AUTHORS McAdam,S.N., Boyson,J.E., Liu,X., Garber,T.L., Hughes,A.L.,  
Bontrop,R.E. and Watkins,D.I.  
TITLE A uniquely high level of recombination at the HLA-B locus  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (13), 5893-5897 (1994)  
MEDLINE 94286544  
REFERENCE 2 (bases 1 to 1065)  
AUTHORS Boyson,J.E.  
TITLE Direct Submission  
JOURNAL Submitted (25-JAN-1994) Jonathan E. Boyson, Watkins Unit, Wisconsin  
Regional Primate Research Center, 1220 Capitol Court, Madison, WI  
53715, USA

FEATURES  
source Location/Qualifiers  
1..1065  
/organism="Pan paniscus"  
/isolate="Bosondjo"  
/db\_xref="taxon:9597"  
1..1065  
/gene="HLA-B"  
<1..1065  
/gene="HLA-B"  
/note="HLA-B-02 allele"  
/codon\_start=1  
/product="class I histocompatibility antigen"  
/protein\_id="AAA50181.1"  
/db\_xref="GI:454774"  
/translation="VLLLSAALALTTWAGSHSMRYFTYSRPGAGPEFTISGVY  
DDTFQVRFSDAASPREAPRMWEGPEYWDNRNTOICKAAQOTDRENLRLALRYN  
OSAGSHTLTQWYGCDDPGDGLRGYRFAYVDGDIYALNEDLSWTAAQTQ  
RKWEARVAEQRAYLEGICVWLRRLYENGKTLORADPPKTHVTHHPSHEATLR  
CWALGYPEITLTWDRGDDQDTDELVELTRPAGDTFQKAAVAVVPGSQORTCH  
VQHEGLPELTLLWEPSSQSTPIVIGIVAGLAVLVVTVGAVVAVMCRKSSGKGG  
SYSQAASSDSQAQSDVSLTA"

BASE COUNT 209 a 327 c 358 g 171 t  
ORIGIN

alignment\_scores  
Quality: 70.00 Length: 15  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 93.333

alignment\_block  
US-08-653-294-33 x PPU05578 ..  
Align seg 1/1 to: PPU05578 from: 1 to: 1065  
1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
:::|||||  
256 CAGGCACAGACTACCGAGAACCTCGGATCGCGTCCGCTAC 300

seq\_name: gb\_pr1:GGB0101

seq\_documentation\_block: 1270 bp mRNA PRI 01-JUL-1992  
LOCUS GGB0101  
DEFINITION G.gorilla MhcGogo-B0101 gene for Mhc class I heavy chain.  
ACCESSION X60255  
VERSION X60255.1 GI:22865  
KEYWORDS B locus allele; Class I major histocompatibility gene; heavy chain;  
major histocompatibility gene; peptide binding protein;  
transplantation antigen.  
gorilla.  
SOURCE  
ORGANISM Gorilla gorilla  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Gorilla.  
1 (bases 1 to 1270)  
AUTHORS Lawlor,D.A.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUN-1991) D.A. Lawlor, Stanford Univ School of

REFERENCE 2 (bases 1 to 1089)  
AUTHORS Lawlor,D.A., Warren,E., Taylor,P. and Parham,P.  
TITLE Gorilla class I major histocompatibility complex alleles:  
comparison to human and chimpanzee class I  
J. Exp. Med. 174 (6), 1491-1509 (1991)  
MEDLINE 92078860  
COMMENT See X60248-59 & X60892-3  
FEATURES  
source Location/Qualifiers  
1..1270  
/organism="Gorilla gorilla"  
/isolate="Banga"  
/db\_xref="taxon:9593"  
/cell\_type="lymphoblast"  
/cell\_line="Banga B lymphoblastoid cell line"  
/clone\_lib="BAB192"  
1..1089  
/gene="MhcGogo-B0101"  
1..73  
/gene="MhcGogo-B0101"  
/number=1  
1..1089  
/gene="MhcGogo-B0101"  
/codon\_start=1  
/product="Mhc class I heavy chain"  
/protein\_id="CAA42807.1"  
/db\_xref="GI:22866"  
/db\_xref="SWISS-PROT:P30379"  
/translation="MRVTAPRTLLLSAALALTTWAGSHSMRYFDFAVRPGRGEP  
RFTIYVDDTQFVRFSDAASPREAPRMWEGPEYWDRETQTSKAQOTDRENL  
RIALYVDDTSAGSHTIQRMFGCDVDPGLLRYSQAYDGKDIALNEDLSWTAA  
DTAAITKRCWAEAREASQRAYLEGTCEVLRRLYENGKTLORADTPKTHVTHHPI  
SDHEATLRCWALGFYPAITLTWDRGDDQDTDELVELTRPAGDTFQKAAVAVVPSG  
EEERTCHVQHEGLPKPLTLRWPSSQSTPIVIGIVAGLAVLVVTVGAVVAVTAVICRR  
KSSGSGSYQAASSDSQAQSDVSLTA"  
1..75  
/gene="MhcGogo-B0101"  
73..1086  
/gene="MhcGogo-B0101"  
/product="Mhc class I heavy chain"  
74..343  
/gene="MhcGogo-B0101"  
/note="alpha 1"  
/number=2  
344..619  
/gene="MhcGogo-B0101"  
/note="alpha 2"  
/number=3  
620..895  
/gene="MhcGogo-B0101"  
/note="alpha 3"  
/number=4  
896..1012  
/gene="MhcGogo-B0101"  
/note="transmembrane"  
/number=5  
1013..1045  
/gene="MhcGogo-B0101"  
/note="cytoplasmic 1"  
/number=6  
1046..1089  
/gene="MhcGogo-B0101"  
/note="cytoplasmic 2"  
/number=7  
1090..>1270  
3'UTR 256 a 388 c 410 g 216 t  
BASE COUNT  
ORIGIN

alignment\_scores  
Quality: 70.00 Length: 15  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 93.333

## alignment\_block:

US-08-653-294-33 x GGB0101

Align seg 1/1 to: GGB0101 from: 1 to: 1270

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
 ::

280 CAGGCACAGACTGACCGAGAGACCTGCGGATCGCGCTCGCTAC 324

seq\_name: gb\_prl:GGB0102

seq\_documentation\_block:

LOCUS GGB0102 1270 bp mRNA PRI 07-FEB-1992  
 DEFINITION G-gorilla MhcGogo-B0102 gene for Mhc class I heavy chain.

ACCESSION X60693

VERSION X60693.1 GI:22867

KEYWORDS B locus allele; Class I major histocompatibility gene; heavy chain;  
 major histocompatibility gene; peptide binding protein;

transplantation antigen.

## SOURCE

gorilla

## ORGANISM

Gorilla gorilla

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Gorilla.

Lawlor,D.A.

Direct Submission

Submitted (25-JUN-1991) D.A. Lawlor, Stanford Univ School of

Medicine, Dept of Cell Biology, Stanford CA 94305, USA

2 (bases 1 to 1089)

Lawlor,D.A., Warren,E., Taylor,P. and Parham,P.

Gorilla class I major histocompatibility complex alleles:

comparison to human and chimpanzee class I

J. Exp. Med. 174 (6), 1491-1509 (1991)

92078860

COMMENT See X60248-59 &amp; X60692-3.

## FEATURES

Location/Qualifiers

1..1270

/organism="Gorilla gorilla"

/isolate="Calabar"

/db\_xref="taxon:9593"

/cell\_type="lymphoblast"

/cell\_line="Calabar B lymphoblastoid cell line"

/clone\_lib="CAB181"

1..1089

/gene="MhcGogo-B0102"

1..73

/gene="MhcGogo-B0102"

/number=1

1..1089

/gene="MhcGogo-B0102"

/codon\_start=1

/product="Mhc class I heavy chain"

/protein\_id="CAA43101.1"

/db\_xref="GI:22868"

/db\_xref="SWISS-PROT:P30380"

/translation="MEVTAPRTLLILLAAALATETWAGSHSMRYEDTAVSRGGEPE

RFITGVYDDQFVRFSDAASPRMEPRAPWIEQGPYWRBTQTSKAQQTRENL

RIAUYNQSEAGSHTFQFMFGCDVDPGRLLRGYSQAYDGDYDIALNEDLSWTAA

DTAAQITQRWEAREAEQLRYEGTCVWLRYLENGRETLDQADTPKTHVTHPI

SDHEATLCWALGFYPAITLTWQDGEDQDTLVETRPAGDTFOKWAIVVPSG

EEERTYCHVQHEGLPKPLTLRWEPSQSTIPVIGVAGLAVLVVVIGAVTVICRR

KSSGKGSGYSQAASDSSAQSDVSLTA"

1..75

/gene="MhcGogo-B0102"

73..1086

/gene="MhcGogo-B0102"

/product="Mhc class I heavy chain"

74..343

/gene="MhcGogo-B0102"

/note="alpha 1"

/number=2

344..619

sig\_peptide

mat\_peptide

exon

exon

/gene="MhcGogo-B0102"

/note="alpha 2"

/number=3

620..895

/gene="MhcGogo-B0102"

/note="alpha 3"

/number=4

896..1012

/gene="MhcGogo-B0102"

/note="transmembrane"

/number=5

1013..1045

/gene="MhcGogo-B0102"

/note="cytoplasmic 1"

/number=6

1046..1089

/gene="MhcGogo-B0102"

/note="cytoplasmic 2"

/number=7

1090..&gt;1270

BASE COUNT 255 a 388 c 410 g 217 t

ORIGIN

## alignment\_scores:

Quality: 70.00 Length: 15

Ratio: 4.667 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 93.333

## alignment\_block:

US-08-653-294-33 x GGB0102

Align seg 1/1 to: GGB0102 from: 1 to: 1270

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15

::

280 CAGGCACAGACTGACCGAGAGACCTGCGGATCGCGCTCGCTAC 324

seq\_name: gb\_prl:GGB0103

seq\_documentation\_block:

LOCUS GGB0103 1270 bp mRNA PRI 07-FEB-1992

DEFINITION G-gorilla MhcGogo-B0103 gene for Mhc class I heavy chain.

ACCESSION X60254

VERSION X60254.1 GI:22869

KEYWORDS B locus allele; Class I major histocompatibility gene; heavy chain;

major histocompatibility gene; peptide binding protein;

transplantation antigen.

SOURCE

gorilla

ORGANISM

Gorilla gorilla

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Gorilla.

Lawlor,D.A.

Direct Submission

Submitted (25-JUN-1991) D.A. Lawlor, Stanford Univ School of

Medicine, Dept of Cell Biology, Stanford CA 94305, USA

2 (bases 1 to 1089)

Lawlor,D.A., Warren,E., Taylor,P. and Parham,P.

Gorilla class I major histocompatibility complex alleles:

comparison to human and chimpanzee class I

J. Exp. Med. 174 (6), 1491-1509 (1991)

92078860

COMMENT See X60248-59 &amp; X60692-3.

FEATURES

Location/Qualifiers

1..1270

/organism="Gorilla gorilla"

/isolate="Banga"

/db\_xref="taxon:9593"

/cell\_type="lymphoblast"

/cell\_line="Banga B lymphoblastoid cell line"

/clone\_lib="BAB197"

1..1089

gene

/gene="MhcGogo-B0103"  
 1. .73  
 /gene="MhcGogo-B0103"  
 /number=1  
 1. .1089  
 /gene="MhcGogo-B0103"  
 /codon\_start=1  
 /product="Mhc class I heavy chain"  
 /protein\_id="CAA42806.1"  
 /db\_xref="GI:22870"  
 /db\_xref="SWISS-PROT:P30381"  
 /translation="MVTAPRTLLLSAALATETWAGSHSMRYFTAVSRPGRGEP  
 RFTIVGVDDTQVRFDSDAASPRMEPRAPWIEQGEYWDRETQTSKAAQOTDRENL  
 RIALRYNDSSEAGSHITQWYCDMGDPDGLRLGYQSADGKDYIALNEDLSWTA  
 DTAQITQRKWEAREAEQLRAYLESTCVENLRYLENGREILQADTPKTHVTHPI  
 SDEATRLWALGFYPAETITLWQDGEDQDTLWTRPAGDGTFFQKAAVVPVPSG  
 EERYTCHVQHEGLPKPLRLWEPSSQSTPIVIGVAGLAVLVAVVIGAVVTAICRR  
 KSSGGKGGYSQAASSDSAQGSDVSLTA"  
 1. .75  
 /gene="MhcGogo-B0103"  
 /number=2  
 344. .619  
 /gene="MhcGogo-B0103"  
 /note="alpha 2"  
 /number=3  
 620. .895  
 /gene="MhcGogo-B0103"  
 /note="alpha 3"  
 /number=4  
 896. .1012  
 /gene="MhcGogo-B0103"  
 /note="transmembrane"  
 /number=5  
 1013. .1045  
 /gene="MhcGogo-B0103"  
 /note="cytoplasmic 1"  
 /number=6  
 1046. .1089  
 /gene="MhcGogo-B0103"  
 /note="cytoplasmic 2"  
 /number=7  
 1090. .>1270

sig\_peptide

mat\_peptide

/gene="MhcGogo-B0103"  
 /product="Mhc class I heavy chain"  
 /number=2  
 /note="alpha 1"  
 /number=2

exon

exon

exon

exon

exon

exon

3'UTR

BASE COUNT 257 a 388 c 409 g 216 t  
 ORIGIN

alignment\_scores:  
 Quality: 70.00 Length: 15  
 Ratio: 4.667 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 93.333  
 alignment\_block:  
 US-08-653-294-33 x GGB0103 ..  
 Align seg 1/1 to: GGB0103 from: 1 to: 1270

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
 :::  
 280 CAGGCACAGACTGACCGAGAGAACTGCGGATCGGCTCCGCTAC 324

seq\_name: gb\_pr2:HLU50091

seq documentation block:

LOCUS HLU50091 1065 bp mRNA PRI 14-SEP-1996  
 DEFINITION Hylobates lar MHC class I (Hyla B\*01) mRNA, partial cds.  
 ACCESSION U50091  
 VERSION U50091.1 GI:1545833

KEYWORDS common gibbon.  
 SOURCE Hylobates lar  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.  
 REFERENCE 1 (bases 1 to 1065)  
 AUTHORS Chen, Z.W., McAdam, S.N., Hughes, A.L., Dogon, A.L., Letvin, N.L. and Watkins, D.I.  
 TITLE Molecular cloning of orangutan and gibbon MHC class I cDNA. The HLA-A and -B loci diverged over 30 million years ago  
 JOURNAL J. Immunol. 148 (8), 2547-2554 (1992)  
 MEDLINE 92218857  
 REFERENCE 2 (bases 1 to 1065)  
 AUTHORS Chen, Z.W., McAdam, S.N., Hughes, A.L. and Watkins, D.I.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-FEB-1996) Pathology, Wisconsin Regional Primate Research Center, 1220 Capitol Court, Madison, WI 53703, USA  
 FEATURES  
 Location/Qualifiers  
 1. .1065  
 /organism="Hylobates lar"  
 /db\_xref="taxon:9580"  
 <1. .1065  
 /note="Hyla B\*01"  
 /codon\_start=1  
 /product="MHC class I"  
 /protein\_id="AA08074.1"  
 /db\_xref="GI:1545834"  
 /translation="LLLLLSAALATETWAGSHSMRYFTSVSRGRGEPRTIVGVY  
 DDTQVRFDSDAASPRMEPRAPWIEQGEYWDQQTQISKNAQTDRENLRIALRYN  
 QSEDSHTIQWYCDGLPDLRLGYQYADGKDYIALNEDLSWTAADTAQITQ  
 RKWEAAHEAEQMRAYLEGRLEWLRYLENGKETHQRAEPPETHVTHPISTDEATLR  
 CWALGFYPAETITLWQDGEDQDTLWTRPAGDGTFFQKAAVVPVPSGEQRYTCH  
 VQHEGLPEPLRLWEPSSQSTPIVIGVAGLAVLVAVVIGAVVAVVMMRRKSLSGGKG  
 SYSKRASSDSAQGSDVSLTA"  
 212 a 329 c 354 g 170 t  
 BASE COUNT  
 ORIGIN  
 alignment\_scores:  
 Quality: 69.00 Length: 14  
 Ratio: 4.929 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-653-294-33 x HLU50091 ..  
 Align seg 1/1 to: HLU50091 from: 1 to: 1065  
 2 AlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
 :::  
 259 GCACAGACTGACCGAGAGAACTGCGGATCGGCTCCGCTAC 300  
 seq\_name: gb\_pri:PTU05580  
 seq documentation block:  
 LOCUS PTU05580 1068 bp mRNA PRI 08-OCT-1994  
 DEFINITION Pan troglodytes class I histocompatibility antigen Patr-B  
 (Patr-B-09 allele) mRNA, partial cds.  
 ACCESSION U05580  
 VERSION U05580.1 GI:454777  
 KEYWORDS chimpanzee.  
 SOURCE Pan troglodytes  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 REFERENCE 1 (bases 1 to 1068)  
 AUTHORS McAdam, S.N., Boyson, J.E., Liu, X., Garber, T.L., Hughes, A.L., Bontrop, R.E. and Watkins, D.I.  
 TITLE A uniquely high level of recombination at the HLA-B locus  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (13), 5893-5897 (1994)  
 MEDLINE 94286544  
 REFERENCE 2 (bases 1 to 1068)  
 AUTHORS Boyson, J.E.



REITVCYVDDTQVFRSDAASPRMEPRAPWLEQEGPEYWDRETRNVKASQTDRENIL  
RIALRYNOSAGSHWTQWYCDVGDGRLRGYQFAYDKDYLALNODSSWTAA  
DYAAQITQRKWAERAEQRAYLEGTCTVEWLRRLLENGKETLQRADPKTHVTHHPHI  
SHEATLRWALGFYPAEITLTWQDGEDOTQDTLVEVTRPAGDRTFKWAAVVPSPG  
EQRVYCHVQHEGLPKPLRLRWPSSQSTIPVIGVAGLAVLVVTVAVVAANVCRKRS  
KSSGGKGGYSQAASDSSAQGSVSLTA"

BASE COUNT 215 a 338 c 363 g 173 t  
ORIGIN

alignment\_scores:  
Quality: 69.00 Length: 14  
Ratio: 4.929 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-33 x AF115460 ..

Align seg 1/1 to: AF115460 from: 1 to: 1089

2 AlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
283 GCACAGACTGACCGAGAGAACCTGCGGATCGCGCTCGCTAC 324

seq\_name: gb\_pr2:PTCHLAB1

seq\_documentation\_block:  
LOCUS PTCHLAB1 1507 bp mRNA PRI 27-MAR-1995  
DEFINITION Chimpanzee mRNA for class I antigen CHLA-B1.  
ACCESSION X13115 X12781

VERSION X13115.1 GI:38205

KEYWORDS antigen; cell surface glycoprotein; class I antigen; major histocompatibility complex; surface antigen.

SOURCE

ORGANISM

Pan troglodytes  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Pan.  
Mayer, W.E.  
1 (bases 1 to 1507)

REFERENCE

AUTHORS

TITLE Direct Submission

JOURNAL Submitted (03-OCT-1988) W.E. Mayer, Stanford University, Dept. of Biological Sciences, Stanford, CA 94305-5020, USA

REFERENCE

AUTHORS

Mayer, W.E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G. and Klein, J.

TITLE Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution

JOURNAL EMBO J. 7 (9), 2765-2774 (1988)

MEDLINE 89030641

REFERENCE

AUTHORS

Mayer, W.

TITLE Direct Submission

JOURNAL Submitted (17-FEB-1989) to the EMBL/GenBank/DBJ databases

COMMENT Position 607 is c in [1], [3] and u in [2].

FEATURES

SOURCE

1. .1507

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/chromosome="5."

/tissue\_type="lymphocyte"

/cell\_line="Hugo-EBV"

/clone="p1B1, p1B4."

<1..62

/product="leader peptide (AA -21 to -1)"

<1..1082

/note="CHLA-B1 protein precursor (1 is 2nd base in codon)

(AA -21 to 339)"

/codon\_start=3

/protein\_id="CAA31507.1"

/db\_xref="GI:755776"

/translation="APRTVLLLSAALATETWAGSHMRVYTSVSRPGRGEPRFIT

VGVDVDTQVFRSDAASPRMEPRAPWLEQEGPEYWDRETRNVKASQTDRENILRIAL

RYVNOSEAGSHWTQWYCDVGDGRLRGYQFAYDKDYLALNEDLSSWTAAADTAA  
QITQRKWAERAEQRAYLEGTCTVEWLRRLLENGKETLQRADPKTHVTHHPISDHE  
ATLRWALGFYPAEITLTWQDGEDOTQDTLVEVTRPAGDRTFKWAAVVPSPGEEQR  
YCHVQHEGLPKPLRLRWPSSQSTIPVIGVAGLAVLVVTVAVVAANVCRKRS  
GKGGYSQAASDSSAQGSVSLTA"

mat\_peptide

/product="CHLA-B1 mature protein (AA 1-339)"

misc\_feature

63..332

/note="alpha-1 domain"

misc\_feature

333..608

/note="alpha-2 domain"

misc\_feature

609..884

/note="alpha-3 domain"

misc\_feature

885..1001

/note="transmembrane domain"

misc\_feature

1002..1079

/note="cytoplasmic domain"

misc\_feature

1486..1491

/note="polyA signal"

polyA\_site

1507

/note="polyA site"

BASE COUNT 318 a 435 c 452 g 302 t

ORIGIN

alignment\_scores:

Quality: 69.00 Length: 14

Ratio: 4.929 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-33 x PTCHLAB1 ..

Align seg 1/1 to: PTCHLAB1 from: 1 to: 1507

2 AlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
273 GCACAGACTGACCGAGAGAACCTGCGGATCGCGCTCGCTAC 314

seq\_name: gb\_pr1:CHPMHCAA

seq\_documentation\_block:

LOCUS CHPMHCAA 1537 bp mRNA PRI 27-SEP-1993

DEFINITION Chimpanzee MHC class I ChLA chain mRNA, complete cds, clone 39.

ACCESSION M24045 M24029

VERSION M24045.1 GI:176814

KEYWORDS cell surface antigen; cell surface glycoprotein; class I gene; integral membrane protein; major histocompatibility complex.

SOURCE Chimpanzee (Tank) EBV-transformed B lymphoblastoid cell, CDNA to mRNA, clone 39.

ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 (bases 1 to 87; 1180 to 1537)

AUTHORS Ennis, P.D.

JOURNAL Unpublished (1989)

AUTHORS Parham, P., Lawlor, D.A., Lomen, C.E. and Ennis, P.D.

TITLE Diversity and diversification of HLA class I alleles

JOURNAL J. Immunol. 142, 3937-3950 (1989)

MEDLINE 89235215

COMMENT Draft entry and computer-readable sequence for [2] kindly submitted by P.D. Ennis, 14-APR-1989.

FEATURES

Location/Qualifiers

1..1537

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/map="5"

<1..1537

/note="MHC ChLA mRNA"

20..1111

/note="MHC ChLA chain"

/codon\_start=1



/protein\_id="AAA35424.1"  
/db\_xref="GI:176815"  
/translation="MQVTAPRTVLLLSAALALTETWAGSHSMRYEDTAVSRPGRGEP  
REITVGVDDTQVFRFSDAASPRMEPRAPWIEQEGPEYWDRETRNVKASQTDRENL  
RIALRYINOSKAGSHTWQFMGCDMGPDGRLRGYQYAYDGDYIALNEDLUSWTAA  
DTAAQITQRKWAARAAEQRAYLEGTCVWLLRRLRYLNGKETLQRADPPKTHMTHPI  
SDHEATLRCAWLGFPYPAEITLTWQDGEDQTDTELVTETPEGDRTFQKAAAVVVP  
EQRYTCHVQHEGLPKPLTLRWEPSQSQTIPVIGIVAGLAVLWTVAVVAVVAVLCR  
RKSSGGKGSYSQAASSDSAQGSYSLTA"  
BASE COUNT 324 a 445 c 459 g 309 t  
ORIGIN

alignment\_scores:  
Quality: 69.00 Length: 14  
Ratio: 4.929 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-33 x CHPMHCAA ..

Align seg 1/1 to: CHPMHCAA from: 1 to: 1537

2 AlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
302 GCACAGACTGACCGAGAGAACCTCGGATCGCGCTCGCTAC 343

**THIS PAGE BLANK (USPTO)**



```

FT intron 3009. 3041
FT /*tag- f
FT intron 3148. 3191
FT /*tag- g
PN EP-226069-A.
PD 24-JUN-1987.
PF 21-NOV-1986; 116139.
PR 01-JAN-1985; DE-542024.
PR 21-DEC-1985; DE-545576.
PA (BEHW ) BEHRINGER AG.
PI Szots H, Weiss E, Dornier C, Lang M, Meo T, Riethmuller G;
DR P-PSDB: P70155.
PT DNA coding for human histocompatibility antigen HLA-B 27 - useful
PT for diagnosis and antigen and antibody prodn.
PS Claim 1; p6; 13pp; German.
CC The DNA may be used to detect the HLA-B 27 gene (opt. mutated) in
CC human genetic material. The HLA-B 27 may be used to detect anti-HLA-
CC B 27 antibodies in human serum. The antibodies may be used to
CC determine HLA-B 27 levels in human serum, e.g. for diagnosis of
CC rheumatic disorders, esp. ankylosing spondylitis.
SQ Sequence 3874 BP; 751 A; 1094 C; 1171 G; 858 T;

alignment_scores:
Quality: 59.00 Length: 15
Ratio: 4.538 Gaps: 0
Percent Similarity: 86.667 Percent Identity: 80.000

alignment_block:
US-08-653-294-33 x N70225 ..
Align seg 1/1 to: N70225 from: 1 to: 3874

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15
|||||
926 AAGGCACAGACTGACCGAGAGGACCTGCGGACCTGCTCCGCTAC 970

seq_name: N_Geneseq_36:T61639

seq_documentation_block:
ID T61639 standard; DNA; 6553 BP.
AC T61639;
DT 05-JUN-1997 (first entry)
DE HLA B27 consensus sequence.
KW HLA B27; seronegative spondylarthropathy; ankylosing spondylitis;
KW Reiter's syndrome; arthritis; acute anterior uveitis; diagnosis;
KW ss; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT mrna 3968. 6653
FT /*tag- a
FT /note- "HLA-B27 3' flanking region, downstream of
FT 3' untranslated region"
FT mrna 4112. 4556
FT /*tag- b
FT /note- "3' flanking region diagnostic for genetic
FT predisposition to SNSA"
FT mrna 4270. 4556
FT /*tag- b
FT /note- "3' flanking region diagnostic for genetic
FT predisposition to SNSA"
FT misc_difference 4495
FT /*tag- d
FT /note- "absence of cytosine at this site is
FT indicative of a predisposition to SNSA"

PN WO9709450-A1.
PD 13-MAR-1997.
PF 16-AUG-1996; U13256.
PR 01-SEP-1995; US-522942.
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PI Tyan DB;
DR WPI; 97-192924/17.

```

```

PT Detecting pre-disposition to seronegative spondylarthropathies -
PT from the absence of a C residue at a specific position in the
PT 3'-flanking region of the HLA B27 allele.
PS Claim 1; Page 52-56; 68pp; English.
CC Genetic predisposition to seronegative spondylarthropathies (SNSA)
CC is detected by determining the absence of a cytosine nucleotide in
CC the 3' flanking region (see also T61647-48) of an HLA-B gene at a
CC position corresponding to nucleotide 4495 of the HLA-B27 consensus
CC sequence given in T61639. Probes and primers (see also T61640-46)
CC based on this region can be used in diagnostic assays to detect the
CC genetic predisposition to SNSA, and permit the distinction of B27+
CC individuals who are resistant to SNSA, from B27+ normal individuals
CC who are susceptible (but as yet unaffected) to such diseases.
SQ Sequence 6553 BP; 1443 A; 1619 C; 2017 G; 1474 T;

alignment_scores:
Quality: 59.00 Length: 15
Ratio: 4.538 Gaps: 0
Percent Similarity: 86.667 Percent Identity: 80.000

alignment_block:
US-08-653-294-33 x T61639 ..
Align seg 1/1 to: T61639 from: 1 to: 6553

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15
|||||
1087 AAGGCACAGACTGACCGAGAGGACCTGCGGACCTGCTCCGCTAC 1131

seq_name: N_Geneseq_36:Q29167

seq_documentation_block:
ID Q29167 standard; DNA; 270 BP.
AC Q29167;
DT 09-MAR-1993 (first entry)
DE HLA-Bw 52 exon 2 alpha-1 domain.
KW Human leukocyte antigen; transgenic; germ cells; somatic cells;
KW expression; ss.
PN J04091731-A.
PD 25-MAR-1992.
PF 03-AUG-1990; 207329.
PR 03-AUG-1990; JP-207329.
PA (OLYU ) OLYMPUS OPTICAL CO.
DR WPI: 92-342893/42.
PT Transgenic non-human mammalian HLA-Bw 52 gene - useful for
PT analysis of expression of gene structure, and prodn. of
PT mouse model of human disease
PS Disclosure: Fig 1; 8pp; Japanese.
CC The sequence shows the exon 2 alpha-1 domain of the human leukocyte
CC antigen-Bw 52 gene. The complete gene may be introduced into non-
CC human mammals, pref. rat or mouse, or their ancestors at the primary
CC developmental biological step via transplantation into the zygote or
CC embryo to generate transgenic non-human mammals incorporating the
CC HLA-Bw 52 gene in both their germ cells and somatic cells. Transgenic
CC non-human mammals contg. HLA-Bw 52 are useful for the analysis of
CC expression of the gene, its structure, and prodn. of mouse models of
CC human disease. See also Q29166-72.
SQ Sequence 270 BP; 59 A; 88 C; 86 G; 37 T;

alignment_scores:
Quality: 56.00 Length: 13
Ratio: 4.667 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 92.308

alignment_block:
US-08-653-294-33 x Q29167 ..
Align seg 1/1 to: Q29167 from: 1 to: 270

3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15
|||||

```

213 CAGACTTACCGAGAGAACTCGGATCGGCTCCGCTAC 251

seq\_name: N\_Geneseq\_36:Q01834

seq\_documentation\_block:

ID Q01834 standard; DNA; 1086 BP.

AC Q01834;

DT 19-MAR-1991 (first entry)

DE Sequence encoding HLA-B51 antigen.

KW Probe; HLA class I DNA; immunogen; ss.

OS Homo sapiens.

FN EP354580-A.

PD 14-FEB-1990.

PF 10-AUG-1989.

PR 11-AUG-1988; JP-200758.

PA (OLYU) Olympus Optical Co., Ltd.

PI Kano K, Takiguchi;

DR WPI; 90-046289/07.

PT New DNA for class I human leucocyte antigens and derived probes and

PT transformed cells, useful for DNA typing, as immunogens etc.

PS Claim 1; Page 11; 23pp; English.

CC The HLA class I DNA can be used as a source of probes for use in DNA

CC typing. Transformed cells, which are useful as immunogens, can be

CC obtained by introducing these DNAs into eucaryotic cells.

SQ Sequence 1086 BP; 224 A; 334 C; 356 G; 172 T;

alignment\_scores:

Quality: 56.00 Length: 13  
Ratio: 4.667 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 92.308

alignment\_block:

US-08-653-294-33 x Q01834

Align seg 1/1 to: Q01834 from: 1 to: 1086

3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15

||||| ||||||| ||||||| ||||||| |||||||

285 CAGACTTACCGAGAGAACTCGGATCGGCTCCGCTAC 323

seq\_name: N\_Geneseq\_36:Q01822

seq\_documentation\_block:

ID Q01822 standard; DNA; 1086 BP.

AC Q01822;

DT 19-MAR-1991 (first entry)

DE Sequence encoding HLA-Bw52 antigen.

KW Probe; HLA class I DNA; immunogen; ss.

OS Homo sapiens.

FN Key

FT cds Location/Qualifiers

FT 1..1086

FT /\*tag= a

PN EP-354580-A.

PD 14-FEB-1990.

PF 10-AUG-1989.

PR 11-AUG-1988; JP-200758.

PA (OLYU) Olympus Optical Co., Ltd.

PI Kano K, Takiguchi;

DR WPI; 90-046289/07.

DR P-PSDB; R03142.

PT New DNA for class I human leucocyte antigens and derived probes and

PT transformed cells, useful for DNA typing, as immunogens etc.

PS Claim 2; pp11-12; 23pp; English.

CC The HLA class I DNA can be used as a source of probes for use in DNA

CC typing. Transformed cells, which are useful as immunogens, can be

CC obtained by introducing these DNAs into eucaryotic cells.

SQ Sequence 1086 BP; 223 A; 335 C; 358 G; 170 T;

alignment\_scores:

Quality: 56.00 Length: 13  
Ratio: 4.667 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 92.308

Percent Similarity: 92.308 Percent Identity: 92.308

alignment\_block:

US-08-653-294-33 x Q01822

Align seg 1/1 to: Q01822 from: 1 to: 1086

3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15

||||| ||||||| ||||||| ||||||| |||||||

285 CAGACTTACCGAGAGAACTCGGATCGGCTCCGCTAC 323

seq\_name: N\_Geneseq\_36:Q05693

seq\_documentation\_block:

ID Q05693 standard; DNA; 1089 BP.

AC Q05693;

DT 03-JAN-1991 (first entry)

DE HLA-B51 gene for production of monoclonal antibodies.

KW Allotype specific monoclonal anti-HLA antibodies; hybridomas;

KW transgenic animals; HLA-B51 gene; ss.

FN Key

FT exon Location/Qualifiers

FT 1..73

FT /\*tag= a

FT /number=1

FT 74..343

FT /\*tag= b

FT /number=2

FT /note="alpha 1-domain"

FT 344..619

FT /\*tag= c

FT /number=3

FT /note="alpha 2-domain"

FT 620..895

FT /\*tag= d

FT /number=4

FT /note="alpha 3-domain"

FT 896..1012

FT /\*tag= e

FT /number=5

FT 1013..1042

FT /\*tag= f

FT /number=6

FT 1043..1089

FT /\*tag= g

FT /number=7

PN EP-38183-A.

PD 22-AUG-1990.

PF 07-FEB-1990; 102424.

PR 08-FEB-1989; JP-029313.

PA (OLYU) OLYMPUS OPTICAL KK.

PI Takiguchi M;

DR WPI; 90-255479/34.

PT Allotype specific monoclonal anti- HLA antibodies prodn. - using

PT hybridomas derived from transgenic animals carrying HLA gene and

PT immunised with HLA antigen of different allotype

PS Disclosure: Fig 1 A-G: 20pp; English.

CC The human HLA-B51 gene was injected into fertilised mouse eggs and

CC then these introduced into the uterus of a pseudo pregnant mouse.

CC The young were tested to ensure incorporation of the gene into the

CC chromosome, and one of them mated 3 times with a normal male to

CC produce 16 young, seven of which carried the HLA-B51 gene.

CC The transgeneic offspring were immunised with HLA-B51 antigen.

CC The spleen lymphocytes were fused with myeloma cells. Hybridomas

CC producing antibodies were selected.

CC See also Q05701.

SQ Sequence 1089 BP; 224 A; 335 C; 357 G; 173 T;

alignment\_scores:

Quality: 56.00 Length: 13  
Ratio: 4.667 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 92.308



CC immobilised on the solid support at one time and irradiation with a  
 CC remove the protecting group and allow the probe to hybridise with a  
 CC complementary sequence. Selective irradiation in defined regions  
 CC will activate only those probes required/chosen to be used in a  
 CC given screening of a sample solution. The invention allows  
 CC simultaneous screenings for many different sequences in a single  
 CC step.  
 SQ Sequence 30 BP; 2 A; 9 C; 9 G; 10 T;

alignment\_scores:  
 Quality: 45.00 Length: 9  
 Ratio: 5.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-653-294-33 x Q51474/rev ..  
 Align seg 1/1 to reverse of: Q51474 from: 1 to: 30

2 AlaGlnThrAspArgGluAsnLeuArg 10  
 |||||  
 30 GCACAGACTGACCGAGAGACCTCGG 4

seq\_name: N\_Geneseq\_36:Q75974

seq\_documentation\_block:  
 ID Q75974 standard; cdna: 4059 BP.  
 AC Q75974;  
 DE 23-AUG-1995 (first entry)  
 DE PHLA-B7 expression vector.  
 KW expression vector; pHLA-B7; heavy human HLA-B7; bicistronic mRNA;  
 KW light beta-2 microglobulin; class I major histocompatibility complex;  
 KW MHC; human leukocyte antigen; HLA; covalently closed circular DNA; ds.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..354  
 FT /tag= a  
 FT /note= "pBR322 backbone contg. bacterial origin of  
 FT replication"  
 FT 355..1170  
 FT /tag= b  
 FT /note= "kanamycin resistance gene open reading frame;  
 FT the gene is taken from the transposable element  
 FT trn903"  
 FT complement (1410..1177)  
 FT /tag= c  
 FT /note= "SV40 polyA signal sequence"  
 FT complement (1412..1560)  
 FT /tag= d  
 FT /note= "SV40 small t intron"  
 FT complement (1561..1794)  
 FT /tag= e  
 FT /note= "3' untranslated region of HLA-B7  
 FT heavy chain mRNA"  
 FT complement (1795..2880)  
 FT /tag= f  
 FT /note= "HLA-B7 open reading frame"  
 FT complement (2886..3415)  
 FT /tag= g  
 FT /note= "Rous sarcoma virus 3' LTR promoter region"  
 FT 3416..4059  
 FT /tag= h  
 FT /note= "pBR322 backbone"  
 FT  
 PN W09429469-A.  
 PD 22-DEC-1994.  
 PF 27-MAY-1994; U06059.  
 PR 07-JUN-1993; US-074344.  
 PA (UNMI) UNIV MICHIGAN.  
 PA (VICA-) VICAL INC.  
 PI Lew D, Marquet M, Nabel EG, Nabel GJ;  
 DR WPI: 95-036494/05.  
 PT New vectors for gene therapy, partic for tumours - comprising

PT genetic material encoding one or more cistron(s) which express  
 PT immunogenic or therapeutic peptide(s)  
 PS Claim 9; Page 42-43; 50pp; English.  
 CC This HLA-B7 antigen encoding plasmid was developed to incorporate many  
 CC advantageous features, eg. the kanamycin resistance gene. The  
 CC eradication of two open reading frames encoding portions of SV40 viral  
 CC proteins lowers the risk of tumorigenicity. The vector may also operate  
 CC as a cassette into which cistrons may be inserted and removed at will  
 CC for the transcription and subsequent translation of peptides of interest.  
 CC The vector is used partic. for the treatment of neoplastic disease,  
 CC eg. melanoma, and provides enhanced gene delivery and expression  
 CC in vivo.  
 SQ Sequence 4059 BP; 975 A; 1051 C; 1033 G; 1000 T;

alignment\_scores:  
 Quality: 41.00 Length: 10  
 Ratio: 4.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 80.000

alignment\_block:  
 US-08-653-294-33 x Q75974/rev ..  
 Align seg 1/1 to reverse of: Q75974 from: 1 to: 4059

1 LysAlaGlnThrAspArgGluAsnLeuArg 10  
 ::|||  
 2604 CAGGCACAGACTGACCGAGAGACCTCGG 2575

seq\_name: N\_Geneseq\_36:Q75973

seq\_documentation\_block:  
 ID Q75973 standard; cdna: 4965 BP.  
 AC Q75973;  
 DE 23-AUG-1995 (first entry)  
 DE PHLA-B7/beta-2 microglobulin expression vector.  
 KW expression vector; pHLA-B7/beta-2 microglobulin; heavy human HLA-B7;  
 KW light beta-2 microglobulin; class I major histocompatibility complex;  
 KW MHC; bicistronic mRNA; human leukocyte antigen; HLA;  
 KW covalently closed circular DNA; ds.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT ltr 1..529  
 FT /tag= a  
 FT /note= "Rous sarcoma virus LTR promoter domain, derived  
 FT for the Schmidt-Rupin strain nucleotides  
 FT 8673-9146. This region also includes a 56 bp  
 FT region of a synthetic oligonucleotide which  
 FT modifies this regulatory sequence to effect a  
 FT higher level of expression of downstream  
 FT sequences. The oligonucleotide removes a  
 FT polyadenylation signal sequence originally found  
 FT in the RSV DNA sequence."  
 FT 531..534  
 FT /tag= b  
 FT /label= consensus\_Kozak\_signal\_sequence  
 FT 535..1620  
 FT /tag= c  
 FT /note= "HLA-B7 heavy chain open reading frame"  
 FT 535..606  
 FT /tag= d  
 FT /note= "encodes putative signal peptide of the  
 FT HLA-B7 heavy chain"  
 FT 607..1620  
 FT /tag= e  
 FT /note= "encodes putative HLA-B7 heavy chain mature  
 FT peptide"  
 FT 1621..1853  
 FT /tag= f  
 FT /note= "3' untranslated sequence of HLA-B7 heavy  
 FT chain mRNA"  
 FT 1854..1888  
 FT /tag= g  
 FT misc\_feature

```

FT FT /note= "multiple cloning site, forms a junction
FT FT between the HLA-B7 sequence and the EMCV-CITE
FT FT sequence, and is used to facilitate subcloning"
FT FT 1889. .2479
FT FT /*tag= h
FT FT /note= "murine encephalomyocarditis CAP-independent
FT FT translational enhancer (EMCV-CITE); taken from
FT FT nucleotides 255-843 of cloned EMCV genomic DNA.
FT FT It is a non-coding regulatory sequence functioning
FT FT as an internal entry point for the eukaryotic
FT FT ribosomal subunits when located within a mRNA
FT FT mol. . It enables the translational start codon of
FT FT the beta-2 microglobulin, downstream of the HLA-B7
FT FT stop codon on this bicistronic mRNA to be
FT FT recognised by the ribosome"
FT FT 2480. .2839
FT FT /*tag= i
FT FT /note= "encodes beta-2 microglobulin; this cDNA is
FT FT deriv. from chimpanzee (differs to the human
FT FT cDNA by only 4 bases)"
FT FT 2840. .2846
FT FT /*tag= j
FT FT /note= "3' untranslated region of the beta-2
FT FT microglobulin mRNA"
FT FT 2847. .2870
FT FT /*tag= k
FT FT /note= "synthetic linker"
FT FT 3'UTR .3111
FT FT /*tag= l
FT FT /note= "bovine growth hormone 3'UTR and transcriptional
FT FT terminator; it starts at a blunt-ended BglI site
FT FT within the 3'UTR of the mRNA coding sequence"
FT FT 2979. .2984
FT FT /*tag= m
FT FT 3112. .3151
FT FT /*tag= n
FT FT /note= "synthetic linker to facilitate cloning"
FT FT complement (3151. .3967)
FT FT /*tag= o
FT FT /note= "kanamycin resistance gene open reading frame;
FT FT the gene is taken from the transposable element
FT FT Tn903"
FT FT 4014. .4965
FT FT /*tag= p
FT FT /note= "pBR322 backbone contg. bacterial origin of
FT FT replication, it represents nucleotides 2244-3193"
FT FT WO9429469-A.
FT FT 22-DEC-1994. PD
FT FT 27-MAY-1994; U06069.
FT FT 07-JUN-1993; US-074344.
FT FT (UNKI ) UNIV MICHIGAN.
FT FT (VICA-) VICAL INC.
FT FT Lew D, Marquet M, Nabel EG, Nabel GJ;
FT FT WPI; 95-036494/05.
FT FT New vectors for gene therapy, partic for tumours - comprising
FT FT genetic material encoding one or more cistron(s) which express
FT FT immunogenic or therapeutic peptide(s)
FT FT Claim 8; Page 41-42; 50pp; English.
FT FT The pHLA-B7/beta-2 microglobulin plasmid expression vector, in addition
FT FT to the kanamycin resistance gene, contains the plasmid DNA encoding the
FT FT heavy (human HLA-B7) and light (beta-2 microglobulin) proteins of a
FT FT class I major histocompatibility complex (MHC) antigen. The plasmid is
FT FT designed to express these two proteins via a bicistronic mRNA in
FT FT eukaryotic cells. Initiation of transcription of the mRNA is dependent
FT FT on a Rous Sarcoma Virus (RSV) promoter sequence deriv. from the 3' long
FT FT terminal repeat. Termination of transcription is dependent upon the
FT FT polyadenylation signal sequence deriv. from the bovine growth hormone
FT FT gene. Eukaryotic cell translation of the heavy chain is regulated by the
FT FT 5' cap-dependent protein start site. Translation of the light chain is
FT FT controlled by the CITE. Finally the replication of the plasmid in
FT FT bacterial cells is controlled by the presence of a bacterial origin of
FT FT replication. The vector is used partic. for the treatment of neoplastic
FT FT disease, eg. melanoma, and provides enhanced gene delivery and expression

```

```

CC in vivo.
SQ Sequence 4965 BP; 1171 A; 1293 C; 1338 G; 1163 T;

alignment_scores:
  Quality: 41.00 Length: 10
  Ratio: 4.100 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:
US-08-653-294-33 x Q75973 ..
Align seg 1/1 to: Q75973 from: 1 to: 4965

1 LysAlaGlnThrAspArgGluAsnLeuArg 10
   ::::::::::::::::::::::::::::
814 CAGGCACAGACTGTACCGAGAGCGTGGG 843

seq_name: N_Geneseq_36:T05843

seq_documentation_block:
ID T05843 standard; DNA; 934 BP.
AC T05843;
DT 03-APR-1996 (first entry)
DE Repeat sequence of cps gene locus operon of S.pneumoniae.
KW Capsular polysaccharide; cps; peptide; flanking region; detection;
KW serotype; diagnosis; prevention; Streptococcus pneumoniae; ss.
OS Streptococcus pneumoniae.
PN WO9531548-A1.
PD 23-NOV-1995.
PF 16-MAY-1995; U061119.
PR 16-MAY-1994; US-243546.
PA (UABR-) UAB RES FOUND.
PI Dillard JP, Yother J;
DR WPI; 96-010934/01.
PT New Streptococcus pneumoniae capsular polysaccharide genes - used
PT for detection, serotyping and for diagnosis and prevention of S.
PT pneumoniae infection
PS Claim 8; Page 154-155; 226pp; English.
CC Sequences encoding the 5' flanking region of the capsular
CC polysaccharide gene (cps) of Streptococcus pneumoniae and which are
CC of sufficient length to allow hybridisation under standard
CC hybridisation conditions to a S.pneumoniae cps gene flanking region
CC may be used in methods to detect and serotype S.pneumoniae. They
CC may also be used for the diagnosis and prevention of S. pneumoniae
CC infection. This sequence corresponds to the downstream region of
CC the 5' cps gene flanking region and is termed the "repeat region".
SQ Sequence 934 BP; 311 A; 136 C; 220 G; 267 T;

alignment_scores:
  Quality: 39.00 Length: 15
  Ratio: 3.545 Gaps: 0
Percent Similarity: 73.333 Percent Identity: 46.667

alignment_block:
US-08-653-294-33 x T05843 ..
Align seg 1/1 to: T05843 from: 1 to: 934

1 LysAlaGlnThrAspArgGluAsnLeuArgLeuAlaLeuArgTyr 15
   ::::::::::::::::::::
79 AAAATCCAGACACCTAAAGCAGACAAAGGGTTGCCATCAGTAT 123

seq_name: N_Geneseq_36:N92449

seq_documentation_block:
ID N92449 standard; DNA; 534 BP.
AC N92449;
DT 08-MAY-1990 (first entry)
DE Sequence of carcinoembryonic antigen domain III.
KW Carcinoembryonic antigen; domain III; domain A; domain B.
FH Key Location/Qualifiers

```



```

FT cds 1..534
FT misc_feature 1..267 a
FT FT /tag= b
FT /note="domain A"
FT misc_feature 268..534
FT FT /tag= c
FT /note="domain B"
PN EP-343946-A.
PD 29-NOV-1989.
PF 24-MAY-1989; 305232.
PR 25-MAY-1988; US-198289.
PA (CITY) City of Hope.
PI Shively JE.
DR WPI; 89-349991/48.
DR P-PSDB; P93499.
PT Carcinoembryonic antigen fragments - used in assays to determine the
PT presence and amt. of the antigen in samples also contg. related antigens.
PS Disclosure; page 4; 15pp; English.
CC CEA fragments can be used in assays to determine the presence and amt. of
CC CEA in samples which also may contain related antigens including its
CC normal cross-reacting antigen or the 128 kD antigen.
SQ Sequence 534 BP; 137 A; 172 C; 110 G; 115 T;

```

## alignment\_scores:

```

Quality: 38.00 Length: 11
Ratio: 4.222 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 63.636

```

## alignment\_block:

```
US-08-653-294-33 x N92449/rev ..
```

```
Align seg 1/1 to reverse of: N92449 from: 1 to: 534
```

```

5 AspArgGluAsnLeuArgIleAlaLeuArgTyr 15
||||:||||| ::|
432 GATAAAGAGAACTGTGTGTGTCGGGTAT 400

```

```
seq_name: N_Geneseq_36:Q71567
```

## seq\_documentation\_block:

```

ID Q71567 standard; DNA; 2031 BP.
AC Q71567;
DT 11-MAY-1995 (first entry)
DE Carcinoembryonic antigen DNA.
KW Carcinoembryonic antigen; CEA; neoplastic diseases; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT mat_peptide 103..2028
FT /tag= a
PN EP-618292-A.
PD 05-OCT-1994.
PF 15-MAR-1994; 103986.
PR 25-MAR-1993; EP-810214.
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI Mach J, Pelegriin A, Tersikh A;
DR WPI; 94-304461/38.
DR P-PSDB; R60619.
PT Carcinoembryonic antigen (CEA) derivs - useful as reagents in
PT immunoassay for diagnosis of neoplastic diseases
PS Claim 4; Page 18; 30pp; English.
CC Q71567 is the DNA sequence which encodes carcinoembryonic antigen (CEA)
CC R60619. CEA is free from cross-reactive CEA-like antigens, it is
CC antigenically indistinguishable from the solution form of CEA shed from
CC tumour cells, and it is devoid of ethanolamine. R60619 can be used in a
CC reagent composition for detecting neoplastic diseases in biological
CC samples, or in an immunoassay process where it can specifically detect
CC the presence of tumour cells in a biological sample e.g. blood.
SQ Sequence 2031 BP; 551 A; 642 C; 417 G; 421 T;

```

## alignment\_scores:

```

Quality: 38.00 Length: 11
Ratio: 4.222 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 63.636
alignment_block:
US-08-653-294-33 x Q71567/rev ..
Align seg 1/1 to reverse of: Q71567 from: 1 to: 2031

```

```

5 AspArgGluAsnLeuArgIleAlaLeuArgTyr 15
||||:||||| ::|
1926 GATAAAGAGAACTGTGTGTGTCGGGTAT 1894

```

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-33 to: EST:\* out\_format : pfs  
Date: Feb 8, 2000 6:23 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q/cgn1/USPFO.spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.2  
-DB=EST -QWMT-rastcap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFM=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-33

Query length: 15

Database: EST:\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 7600.090000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est10:AA151891	+	66.00	233.50	0.0003	255
gb_est26:AA1359260	-	66.00	225.46	0.0009	618
gb_est10:AA147151	+	62.00	211.41	0.0052	581
gb_est38:AL036690	+	59.00	201.57	0.0051	171
gb_est2:R13904	+	59.00	202.62	0.0161	459
gb_est8:C03945	+	56.00	197.84	0.0297	232
gb_est11:AA263158	+	56.00	196.04	0.0374	283
gb_est6:D82221	+	56.00	193.48	0.0519	375
gb_est31:AI096864	+	51.00	174.32	0.6058	748
gb_est23:AI124815	+	50.00	170.65	0.9704	415
gb_est2:R82189	+	49.00	164.16	2.23	567
gb_est22:R20954	+	46.00	158.26	4.75	325
gb_est22:AI007865	-	45.00	152.50	9.95	410
gb_est6:D82177	+	44.00	150.82	12.34	330
gb_est1:AA0333983	-	43.00	147.11	19.86	332
gb_gss10:AA0224339	-	43.00	144.15	29.03	460
gb_gss14:AA0572652	-	43.00	142.81	34.47	533
gb_gss8:AA056225	-	43.00	140.71	45.16	672
gb_gss3:R21407	-	43.00	138.94	56.62	816
gb_gss5:AA053331	+	42.00	139.93	49.92	490
gb_est26:AA817822	+	42.00	139.04	55.91	540
gb_est37:AI943589	+	42.00	136.82	74.39	690
gb_est12:AA319533	+	41.00	150.44	12.95	103
gb_est32:AV058574	+	41.00	144.93	26.28	189
gb_est33:AV104196	+	41.00	140.76	44.85	299
gb_est31:AA0065386	-	41.00	139.28	54.24	332
gb_est32:AA0070735	-	41.00	138.83	57.48	370
gb_est24:C99692	-	41.00	138.14	62.77	399
gb_est1:D25113	+	41.00	137.36	69.41	435
gb_est12:AA310808	+	41.00	137.19	70.90	443
gb_gss4:AA0704919	-	41.00	137.05	72.21	450
gb_est9:C18310	+	41.00	136.56	76.91	475
gb_est26:AI008245	+	41.00	136.40	78.42	483
gb_gss6:AA0828908	-	41.00	136.29	79.55	489
gb_est30:AA057732	-	41.00	136.20	80.50	494
gb_gss9:AA164656	-	41.00	135.70	85.84	522
gb_est32:AI731334	+	41.00	134.56	99.40	592
gb_est32:AI730993	+	41.00	133.53	113.43	663
gb_est24:C99613	+	41.00	133.02	121.04	701
gb_gss11:AA032864	+	41.00	132.49	129.53	743
gb_gss14:AA0579309	-	41.00	130.17	174.39	959

gb\_est1:T24384 - 40.00 135.65 86.35 351 ! T24384 crs1519 lambdazapst R  
gb\_est5:D74882 + 40.00 135.42 88.94 360 ! D74882 CELK089YF Yuj1 Kohar  
gb\_est36:AV202520 - 40.00 135.42 88.94 360 ! AV202520 AV202520 Yuj1 Kohar

seq\_name: gb\_est10:AA151891

seq\_documentation\_block: 255 bp mRNA EST 10-DEC-1996  
LOCUS AA151891  
DEFINITION zo01f06.r1 Stratagene colon (#937204) Homo sapiens cDNA clone  
IMAGE:566435 5' similar to gb:M13497\_cds1 HLA CLASS I  
HISTOCOMPATIBILITY ANTIGEN, A-24(A-9) A\*2401 (HUMAN); mRNA  
sequence.

ACCESSION AA151891

VERSION AA151891.1 GI:1720754

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 255)

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

COMMENT

On May 8, 1995 this sequence version replaced gi:800234.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LNL: contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Trace considered overall poor quality

seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 1.

FEATURES

Location/Qualifiers

1..255

/organism="Homo sapiens"

/db\_xref="GDB:459088"

/db\_xref="taxon:9606"

/clone="IMAGE:566435"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: colon; Vector: Bluescript SK-; Site:1:

Ecoti; Site:2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. T-84 colonic epithelial cell line. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTT 3"

BASE COUNT 57 a 70 c 75 g 44 t 9 others

ORIGIN

alignment\_scores:

Quality: 65.00 Length: 14

Ratio: 4.714 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 92.857

alignment\_block:

US-08-653-294-33 x AA151891 ..

Align seg 1/1 to: AA151891 from: 1 to: 255

2 AlacInThAsPargluAnLeuArgilleAlaLeuArgTyr 15

.....

65 TCACACACTGACCGAGAGACCTCGCGATCGCNCCTCCGTAC 106

```

331 TCACAGACTGACCGAGAG
seq_name: qb_est10:AA147151

```

Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1875  
High quality sequence stops: 384 Source: IMAGE Consortium, LML  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1875 Std Error: 0.00  
Seq primer: M13Rp1  
High quality sequence stop: 384.

FEATURES  
source Location/Qualifiers  
1..459  
/organism="Homo sapiens"  
/db\_xref="GDB:399148"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:26801"  
/clone\_lib="Soares infant brain 1N1B"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="PH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: Lafmid BA; Site:1; Not  
I: Site:2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5',  
ACTGAAGATTCGGCCGCCGCGAATTTTCTTTTTTTTTTTT 3'];  
double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the Lafmid BA vector.  
Library went through one round of normalization. Library  
constructed by Bencio Soares and M.Patima Donaldo."

BASE COUNT 88 a 144 c 146 g 75 t 6 others

ORIGIN

alignment\_scores:  
Quality: 59.00 Length: 13  
Ratio: 4.538 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 92.308

alignment\_block:  
US-08-653-294-33 x R13904 ..

Align seg 1/1 to: R13904 from: 1 to: 459

seq\_name: gb\_est8:C03945

seq\_documentation\_block:  
LOCUS C03945 232 bp mRNA EST 30-JUL-1996  
DEFINITION C03945 Human heart cDNA (YNakamura) Homo sapiens CDNA clone  
3NH02454, mRNA sequence.  
ACCESSION C03945  
VERSION C03945.1 GI:1467196  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 232)  
AUTHORS Tanaka,T., Ogiwara,A., Uchiyama,I., Takagi,T., Yazaki,Y. and  
Nakamura,Y.  
TITLE Construction of a normalized directionally cloned cDNA library from  
adult heart and analysis of 3040 clones by partial sequencing  
Genomics 35 (1), 231-235 (1996)  
JOURNAL MEDLINE 96299762

```

On Oct 24, 1995 this sequence version replaced gi:1040105.
Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp.

FEATURES             source
Location/Qualifiers
  1..232
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="3NHC2454"
    /clone_lib="Human heart cDNA (Ynakamura)"
    /dev_stage="adult"
    /note="Organ: heart; normalized directionally cloned cDNA
          from adult heart"

BASE COUNT      55 a      77 c      68 g      32 t
ORIGIN

alignment_scores:
  Quality:      56.00      Length:      13
  Ratio:        4.667      Gaps:      0
  Percent Similarity: 92.308      Percent Identity: 92.308

alignment_block:
  US-08-653-294-33 x C03945      ..

  Align seg 1/1 to: C03945 from: 1 to: 232

      3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15
      |||||  |||||  |||||  |||||  |||||  |||||
      31 CAGACTTACCGAGAGAACCTGCGGATCGCGCTCCGCTAC 69

seq_name: gb_est11:AA263158

seq_documentation_block:
LOCUS      AA263158      283 bp      mRNA      EST      02-JUL-1998
DEFINITION      PMY0534 KGI-a Lambda Zap Express cDNA library Homo sapiens cDNA 5'
                mRNA sequence.
ACCESSION      AA263158
VERSION        AA263158.1      GI:1898964
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1      (bases 1 to 283)
AUTHORS        Claudio,J.O., Liew,C.C., Dempsey,A.A., Cukerman,E., Stewart,A.K.,
                Na,E., Atkins,H.I., Iscove,N.N. and Hawley,R.G.
TITLE          Identification of sequence-tagged transcripts differentially
                expressed within the human hematopoietic hierarchy
JOURNAL        Genomics 50 (1), 44-52 (1998)
MEDLINE        98292493
COMMENT        On May 5, 1995 this sequence version replaced gi:797810.
                Contact: Hawley RG
                Oncology Research Laboratories
                The Toronto Hospital
                CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
                Tel: 416 3403834
                Fax: 416 3403453
                Email: r.hawley@utoronto.ca
                Similar to M58636 MHC class I HLA-Bw gene. Clone was randomly
                picked from KGIa primary library.
                Seq primer: 5' GAATTAACCTCCTACTAAAGG 3'
                High quality sequence stop: 283.
                Location/Qualifiers
                  1..283
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="KGI-a Lambda Zap Express cDNA library"
                    /cell_type="promyeloblast"

FEATURES             source
Location/Qualifiers
  1..283
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="KGI-a Lambda Zap Express cDNA library"
    /cell_type="promyeloblast"

```

3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
297 CAGACTTACCGAGAGACCTCGGATCGCGTCGCTAC 335

seq\_name: gb\_est31:AI696864

seq\_documentation\_block: 748 bp mRNA EST 03-JUN-1999  
LOCUS AI696864  
DEFINITION WC74h11.x1 NCI-CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2324421 3'  
Similar to gb:M28205 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
B-51(B-5) B\*5101 ALPHA (HUMAN);, mRNA sequence.  
ACCESSION AI696864  
VERSION AI696864.1 GI:4984764  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 748)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 16, 1998 this sequence version replaced gi:2961758.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 424.  
FEATURES  
source  
1..748  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2324421"  
/clone\_lib="NCI-CGAP\_Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life technologies catalog #:  
11548-013"  
BASE COUNT 169 a 227 c 237 g 108 t 7 others  
ORIGIN

alignment\_scores:  
Quality: 56.00 Length: 13  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 92.308 Percent Identity: 92.308

alignment\_block:  
US-08-653-294-33 x AI696864 ..  
Align seg 1/1 to: AI696864 from: 1 to: 748

3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
128 CAGACTTACCGAGAGACCTCGGATCGCGTCGCTAC 166

seq\_name: gb\_est23:AI124815

seq\_documentation\_block: 414 bp mRNA EST 11-SEP-1998  
LOCUS AI124815  
DEFINITION am56e06.x1 Johnston frontal cortex Homo sapiens cDNA clone  
IMAGE:1539586 3' similar to gb:M24038.cdsl HLA CLASS I  
HISTOCOMPATIBILITY ANTIGEN, BW-44(B-12) B\*4402 (HUMAN); contains  
MER22.t3 TARI repetitive element ;, mRNA sequence.

ACCESSION AI124815  
VERSION AI124815.1 GI:3593329  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 414)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Rucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)

COMMENT On Jan 17, 1998 this sequence version replaced gi:1899887.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES  
source  
1..414  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1539586"  
/clone\_lib="Johnston frontal cortex"  
/sex="male"  
/tissue\_type="pooled frontal lobe"  
/dev\_stage="adult"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: Bluescript SK-; Site 1:  
EcoRI; Stanley Neuropathology Consortium  
(www.stanleylab.org) brains S-58, S-65, S-67, S-78.  
Random & oligo-dT primed into EcoRI site of ZAP II Vector.  
Mass excised. Avg insert length 1.9kb. Custom library  
provided by Dr. Nancy Johnston [(410) 614-3918,  
nlj@wellchlink.welch.jhu.edu].  
BASE COUNT 80 a 140 c 136 g 58 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 13  
Ratio: 4.636 Gaps: 0  
Percent Similarity: 84.615 Percent Identity: 84.615

alignment\_block:  
US-08-653-294-33 x AI124815 ..  
Align seg 1/1 to: AI124815 from: 1 to: 414  
3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||  
275 CAGACTTACCGAGAGACCTCGGATCGCGTCGCTAC 313

seq\_name: gb\_est6:D82189

seq\_documentation\_block: 415 bp mRNA EST 09-FEB-1996  
LOCUS D82189  
DEFINITION HUMHBC4524 Human pancreatic islet Homo sapiens cDNA similar to  
HLA-B, mRNA sequence.

ACCESSION D82189  
VERSION D82189.1 GI:1183662  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 415)

```

High quality sequence stop: 345.
Location/Qualifiers
1. 567
/organism="Homo sapiens"
/db_xref="GDB:403812"
/db_xref="taxon:9606"
/clone="IMAGE:31465"
/clone_lib="Soares infant brain lN1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="organ: whole brain; Vector: Lfamid BA; Site1: Not
I; Site2: Hind III; last strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AACTCGAAGATTCGGCCGCGAGGAATTTTTTTTTTTT 3']";
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
107 a 160 c 184 g 107 t 9 others

```

```

alignment_scores:      13
Quality: 50.00      Length:
Ratio: 4.545      Gaps:
Percent Similarity: 84.615      Percent Identity: 84.615

alignment_block:
US-08-653-294-33 x D82189      ..      Align seg 1/1 to: D82189 from: 1 to: 415

3 GlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15
||||| ||||||||| |||||||||
270 CAGACTTACGAGAGACCTTGGATCGCGCTCGGTAC 308

seq name: qb_est2:R20954

```

```

seq_documentation_block: 325 bp mRNA EST 25-JAN-1999
LOCUS AI007865 EST020316 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
DEFINITION RBRAV89 3' end, mRNA sequence.
ACCESSION AI007865
VERSION AI007865.1 GI:3221697
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 325)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900396.
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
            Location/Qualifiers
                1..325
                /organism="Rattus sp."
                /db_xref="ATCC (inhost):2016137"
                /db_xref="taxon:10118"
                /clone="RBRAV89"
                /clone_lib="Normalized rat brain, Bento Soares"

```



/note="Organ: brain; Vector: pT7T3Pac; Site\_1: EcoRI;

BASE COUNT 107 a 61 c 76 g 81 t  
ORIGIN

alignment\_scores:  
Quality: 46.00 Length: 11  
Ratio: 4.182 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 72.727

alignment\_block:

US-08-653-294-33 x AI007865 ..

Align seg 1/1 to: AI007865 from: 1 to: 325

5 AspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||:|||||:|||||:|||||:|||||  
204 GACAGGAAGAATGTACGGTTGGCATTGAGATAC 236

seq\_name: gb\_gss8:AQ033523

seq\_documentation\_block:

LOCUS AQ033523 410 bp DNA GSS 02-JUL-1998  
DEFINITION HS-2229 A2.D07.MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2229 Col-14 Row-G, genomic survey sequence.

ACCESSION AQ033523  
VERSION AQ033523.1 GI:3285641  
KEYWORDS GSS.  
SOURCE human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

#### REFERENCE

1 (bases 1 to 410)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

#### MEDLINE

99380589

#### COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2229 Row: G Column: 14

Class: BAC ends

High quality sequence stop: 410.

Location/Qualifiers

#### FEATURES

##### source

1..410

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-2229 Col-14 Row-G"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 139 a 78 c 62 g 130 t 1 others

#### ORIGIN

alignment\_scores:

Quality: 45.00 Length: 15  
Ratio: 3.214 Gaps: 0  
Percent Similarity: 93.333 Percent Identity: 53.333

alignment\_block:

US-08-653-294-33 x AQ033523/rev ..

Align seg 1/1 to reverse of: AQ033523 from: 1 to: 410

1 LysAlaGlnThrAspArgGluAsnLeuArgIleAlaLeuArgTyr 15  
|||||:|||||:|||||:|||||:|||||  
218 AAGAAGAAGACTGAGACACAGAAATCTAAATAGTTTGGAGGTAT 174

seq\_name: gb\_est6:D82177

seq\_documentation\_block:

LOCUS D82177 330 bp mRNA EST 09-FEB-1996  
DEFINITION HUMHBC4504 Human pancreatic islet Homo sapiens cDNA similar to HLA-B, mRNA sequence.

ACCESSION D82177

VERSION D82177.1 GI:1183645

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 330)

Takeda,J.

Human pancreatic islet ESTs

Unpublished (1995)

On Apr 14, 1993 this sequence version replaced gi:785191.

Contact: Jun Takeda

Institute for Molecular and Cellular Regulation, Gunma University  
3-39-15 Showa-machi, Maebashi Gunma 371, Japan

Tel: 272-20-8856

Fax: 272-20-8896

Email: jtakeda@sp.gunma-u.ac.jp.

#### FEATURES

##### source

1..330

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Human pancreatic islet"

/note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho I; mRNA was prepared from normal adult human islets. cDNA was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."

BASE COUNT 62 a 109 c 106 g 49 t 4 others

#### ORIGIN

alignment\_scores:

Quality: 44.00 Length: 11  
Ratio: 4.400 Gaps: 0  
Percent Similarity: 90.909 Percent Identity: 90.909

alignment\_block:

US-08-653-294-33 x D82177 ..

Align seg 1/1 to: D82177 from: 1 to: 330

3 GlnThrAspArgGluAsnLeuArgIleAlaLeu 13

|||||:|||||:|||||:|||||:|||||

297 CAGACTACCGAGAGAACCTCGGATCGCGTC 329

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 04:05:44 ; Search time 133.56 Seconds  
(without alignments)  
1.773 Million cell updates/sec

Title: US-08-653-294-34  
Perfect score: 51  
Sequence: 1 RESLNRLRGY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	51	100.0	10	R41209	Peptide fragment o
2	51	100.0	10	R83061	HLA-B7 CTL modul
3	51	100.0	10	W07515	T-cell modulating
4	51	100.0	10	W33786	Peptide B7-75-84 t
5	51	100.0	10	W33796	Peptide B7-75-84 t
6	51	100.0	20	R92913	HLA-B7 CTL modul
7	51	100.0	20	R35415	HLA-B7.84-75-84 Pa
8	51	100.0	20	W33790	Peptide B7.84-75/7
9	51	100.0	20	W33797	Peptide fragment o
10	51	100.0	25	R41207	Peptide B7.84-75/7
11	51	100.0	25	R83073	HLA-Bw62 CTL modul
12	51	100.0	25	R35431	HLA-B7.60-84, Comp
13	51	100.0	25	R35419	HLA-Bw62.60-84. Co
14	51	100.0	362	R12464	HLA-B35 antigen. H
15	44	86.3	25	R41206	Peptide fragment o
16	44	86.3	25	R83072	HLA-Bw46 CTL modul
17	44	86.3	25	R95418	HLA-Bw46.60-84. Co
18	44	86.3	366	R12465	HLA-C exon Cb-1. H
19	44	86.3	366	R12466	HLA-C exon Cb-2. H
20	44	86.3	366	Y07033	Breast cancer asso
21	41	80.4	274	R80911	Consensus sequence
22	37	72.5	10	W07518	T-cell modulating
23	35	68.6	22	R44112	HLA epitope. HIV G
24	35	68.6	45	R71629	HLA-alpha-1. Use o
25	35	68.6	489	R47268	Partial sequence o
26	35	68.6	489	R74305	SABP encoded by PC
27	35	68.6	489	W04316	Tobacco salicylic
28	33	64.7	793	R85199	Avenacinase-like p
29	33	64.7	850	W57445	A. thermophilum th
30	33	64.7	1130	W05178	H-lats large tumou
31	33	64.7	1201	W90345	Drosophila sp. Cos
32	32	62.7	9	R42577	HLA epitope. HIV G
33	32	62.7	425	R11329	Alpha subunit of M
34	32	62.7	624	R77674	Glucosylase from

## RESULT 1

R41209  
ID R41209 standard; peptide; 10 AA.  
AC R41209;  
DT 15-MAR-1994 (first entry)  
DE Peptide fragment of Class I HLA peptide.  
KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
KW parasitic disease; cytotoxic T lymphocyte; modulation.  
OS Synthetic.  
PN W09317699-A.  
PD 16-SEP-1993.  
PF 25-FEB-1993; U01758.  
PR 02-MAR-1992; US-844716.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger CA, Krensky AM;  
DR WPI; 93-303134/38  
PT New peptide(s) based on Class I HLA antigen domains - used for  
PT modulating cytotoxic T-lymphocyte activity towards targets  
PS Claim 11; Page 54; 61pp; English.  
CC The peptide is used to modulate cytotoxic T-lymphocyte (CTL)  
CC activity, either by inhibition or stimulation. It can be used  
CC for inhibiting CTL toxicity in transplantations, for inducing CTL  
CC activity in parasitic diseases and neoplasia and in studies on viral  
CC infection. The peptide can also be used for identifying CTLs which  
CC bind to it and removing subsets of CTLs from a T-cell composition.  
CC This peptide sequence is more commonly found within larger peptide  
CC compounds of not more than 30 amino acids in length.  
SQ Sequence 10 AA;

Query Match 100.0%; Score 51; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00074;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLRGY 10  
Db 1 RESLNRLRGY 10

## RESULT 2

R83061  
ID R83061 standard; peptide; 10 AA.  
AC R83061;  
DT 15-MAY-1996 (first entry)  
DE HLA-B7 CTL modulating peptide (B7.75-84).  
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
KW class I MHC; HLA-B7.  
OS Synthetic.  
PN W09526979-A1.  
PD 12-OCT-1995.  
PF 05-APR-1995; U04349.  
PR 05-APR-1994; US-222851.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM, Farham P;  
DR WPI; 95-358582/46.  
PT Extension of acceptance period of transplants from MHC unmatched

35 32 62.7 685 1 W31274  
36 32 62.7 785 1 R04907  
37 32 62.7 854 1 P70347  
38 32 62.7 865 1 R20670  
39 32 62.7 865 1 R23797  
40 32 62.7 868 1 R22615  
41 31 60.8 10 R41211  
42 31 60.8 10 W07521  
43 31 60.8 93 W30133  
44 31 60.8 212 R54922  
45 31 60.8 286 W54071

## ALIGNMENTS

Mouse frizzled-8 p  
Type I interferon  
Varicella-zoster v  
Lipoxygenase. Prod  
Rice lipoxygenase.  
Varicella-zoster v  
Peptide fragment o  
T-cell modulating  
ZAP-70 N-terminal  
Cdl1 polypeptide.  
Banana bunchy top

PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Claim 13; Page 66; 80pp; English.  
 CC This sequence represents a fragment of a class I major histocompatibility  
 CC complex (MHC) antigen. This sequence corresponds to residues 75-84 of  
 CC the alpha-1 domain of the class I MHC HLA-B\*7. This sequence, and the  
 CC peptide fragments represented by R83062-R83085, R83090-R83096, and  
 CC R82907-R82913 can be used to extend the period of acceptance by a  
 CC recipient of a transplant from an MHC unmatched donor. The peptides are  
 CC administered to a patient in conjunction with a subtherapeutic amount of  
 CC an immunosuppressant. This is administered to the patient for a limited  
 CC period of time (compared to the lifetime administration for current  
 CC treatments). The peptides particularly modulate (or inhibit) the  
 CC activity of the cytotoxic T lymphocytes (CTLs) of the patient.  
 CC Sequence 10 AA;  
 SQ

Query Match 100.0%; Score 51; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00074;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 1 RESLNLRGY 10  
 RESULT 3  
 W07515  
 ID W07515 standard; peptide; 10 AA.  
 AC W07515;  
 DT 04-AUG-1997 (first entry)  
 DE T-cell modulating peptide #4.  
 KW mammal; major histocompatibility complex; MHC class I; antigen; perforin;  
 KW insulin-dependent diabetes mellitus; multiple sclerosis; inflammation;  
 KW rheumatoid arthritis; psoriasis; pemphigus vulgaris; Sjogren's disease;  
 KW thyroid disease; Hashimoto's thyroiditis; myasthenia gravis; granzyme;  
 KW autologous target cell; cytokine release; T cell activation; therapy.  
 OS Synthetic.  
 PN W09635443-A1.  
 PD 14-NOV-1996.  
 PF 05-APR-1996; U04710.  
 PR 12-MAY-1995; US-440504.  
 PA (SANG-) SANGSTAT MEDICAL CORP.  
 PI Buelow R;  
 DR WPI: 96-518410/51.  
 PT Treatment of auto-immune disease by admin. of peptide(s) corresp. to  
 PT major histocompatibility complex antigens - esp. for delaying onset  
 PT of clinical symptoms of insulin dependent diabetes by modulating T  
 PT cell mediated attack on target cells  
 PS Claim 7; Page 20; 24pp; English.  
 CC W07512-W07518 represent T-cell modulating peptides that can be used in  
 CC the method of the invention. These sequences are based on a portion of  
 CC the generic peptide corresponding to residues 70-91 of the alpha-1 domain  
 CC of the major histocompatibility complex (MHC) class I antigen (see  
 CC W07510). The method is for affecting the course of an autoimmune disease  
 CC involving T-cell mediated destruction of tissue in mammals. These  
 CC peptides are used especially to treat insulin-dependent diabetes  
 CC mellitus, preferably being administered during the pre-clinical stage to  
 CC delay onset of the disease. Other diseases that can be treated are  
 CC multiple sclerosis, rheumatoid arthritis, psoriasis, pemphigus vulgaris,  
 CC Sjogren's disease, thyroid disease, Hashimoto's thyroiditis, myasthenia  
 CC gravis, etc. The peptides modulate T-cell mediated attack on autologous  
 CC target cells, and may also reduce inflammation, swelling, and release of  
 CC cytokines, perforins, granzymes etc. associated with T cell activation.  
 CC Sequence 10 AA;  
 SQ

Query Match 100.0%; Score 51; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00074;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 1 RESLNLRGY 10  
 RESULT 5  
 W33796  
 ID W33796 standard; peptide; 10 AA.  
 AC W33796;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 PN W09744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.

DB 1 RESLNLRGY 10  
 RESULT 4  
 W33786  
 ID W33786 standard; peptide; 10 AA.  
 AC W33786;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 PN W09744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Buelow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (aa77-76R); aa76 - E or V; aa77 -  
 CC D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84 - a hydrophobic or  
 CC small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha-1 domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 CC Sequence 10 AA;  
 SQ

Query Match 100.0%; Score 51; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00074;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 1 RESLNLRGY 10  
 RESULT 5  
 W33796  
 ID W33796 standard; peptide; 10 AA.  
 AC W33796;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 PN W09744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.

PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI; 98-086530/08;  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 CC Sequence 10 AA;

Query Match 100.0%; Score 51; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00074;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RESLNLRGY 10  
 Db 1 RESLNLRGY 10  
 |||||  
 RESULT 6  
 R92913  
 ID R92913 standard; peptide; 20 AA.  
 AC R92913;  
 DE 16-MAY-1996 (first entry)  
 DE HLA-B\*77:01 CTL modulating peptide (B7.84-75/75-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW Class I MHC; HLA-B\*7.  
 OS Synthetic.  
 PN WO9526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD.) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 15; Page 36; 80pp; English.  
 PS R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
 CC Class I major histocompatibility complex (MHC) antigens. This sequence  
 CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
 CC I MHC HLA-B\*7. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with a  
 CC subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 CC Sequence 20 AA;

Query Match 100.0%; Score 51; DB 1; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RESLNLRGY 10  
 Db 11 RESLNLRGY 20  
 |||||  
 RESULT 7  
 R95415  
 ID R95415 standard; peptide; 20 AA.  
 AC R95415;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B\*78:01-84 (Palindromic)  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD.) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI; 95-194027/25.  
 PT Compens. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 18; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. These sequences can be used to isolate  
 CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane  
 CC protein associated with T-cell activation in mammalian T-cells, and is  
 CC also immunologically cross reactive with the heat shock protein Hsc70.  
 CC p74 is found in a limited number of cell types, but is particularly  
 CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
 CC cell with an amphoteric detergent, and then passed through an affinity  
 CC column containing a covalently bound HLA-B\*702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B\*702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 CC Sequence 20 AA;

Query Match 100.0%; Score 51; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 Db 11 RESLNLRGY 20  
 |||||  
 RESULT 8  
 W33790  
 ID W33790 standard; peptide; 20 AA.  
 AC W33790;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.84-75/75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.

```

PD 27-NOV-1997; U08689.
PF 22-MAY-1997; US-653294.
PR (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise
CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC undesirably attacking cells in a host or in vitro. They can also be
CC used in combination with antigenic peptides or proteins of interest to
CC activate CTLs. They can also inhibit the proliferation of T cells in
CC response to anti-CD3. The peptide can be used for preventing rejection
CC of transplants or for treating autoimmune diseases, e.g. diabetes,
CC rheumatoid arthritis and lupus erythematosus. The products can also be
CC used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 100.0%; Score 51; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLRGY 10
Db 11 RESLNRLRGY 20
|||||
11 RESLNRLRGY 20

RESULT 10
R41207
ID R41207 standard; peptide; 25 AA.
AC R41207;
DT 15-MAR-1994 (first entry)
DE Peptide fragment of Class I HLA peptide.
KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;
KW parasitic disease; cytotoxic T lymphocyte; modulation.
OS Synthetic.
PN WO9317699-A.
PD 16-SEP-1993.
PF 25-FEB-1993; U01758.
PR 02-MAR-1992; US-844716.
PI (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Clayberger CA, Krensky AM;
DR WPI; 93-303134/38.
PT New peptide(s) based on Class I HLA antigen domains - used for
PT modulating cytotoxic T-lymphocyte activity towards targets
PS Claim 10; Page 54; 61pp; English.
CC The peptide (or a fragment of at least 10 amino acids, joined at at
CC least one terminus to a sequence other than that of wild type HLA
CC antigen) is used to modulate cytotoxic T-lymphocyte (CTL) activity,
CC either by inhibition or stimulation. It can be used for
CC inhibiting CTL toxicity in transplantations, for inducing CTL
CC activity in parasitic diseases and neoplasia and in studies on viral
CC infection. The peptide can also be used for identifying CTLs which
CC bind to it and removing subsets of CTLs from a T-cell composition.
SQ Sequence 25 AA;

Query Match 100.0%; Score 51; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLRGY 10
Db 16 RESLNRLRGY 25
|||||
16 RESLNRLRGY 25

RESULT 11
R83073
ID R83073 standard; peptide; 25 AA.
AC R83073;
DT 16-MAY-1996 (first entry)
DE HLA-Bw62 CTL modulating peptide (Bw62.60-84).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW Class I MHC; HLA-Bw62.
OS Synthetic.
PN WO9526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.

```

```

PD 27-NOV-1997; U08689.
PF 22-MAY-1997; US-653294.
PR (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise
CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC undesirably attacking cells in a host or in vitro. They can also be
CC used in combination with antigenic peptides or proteins of interest to
CC activate CTLs. They can also inhibit the proliferation of T cells in
CC response to anti-CD3. The peptide can be used for preventing rejection
CC of transplants or for treating autoimmune diseases, e.g. diabetes,
CC rheumatoid arthritis and lupus erythematosus. The products can also be
CC used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 100.0%; Score 51; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLRGY 10
Db 11 RESLNRLRGY 20
|||||
11 RESLNRLRGY 20

RESULT 9
W33797
ID W33797 standard; peptide; 20 AA.
AC W33797;
DT 19-JUN-1998 (first entry)
DE Peptide B7.84-75/75-84 tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN WO9744351-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PI (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise

```

PR 05-APR-1994: US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI: 95-35862/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 13: Page 32: 80pp; English.  
 CC R3061-R83085, R83090-R83096, and R92907-R92913 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC corresponds to residues 60-84 of the alpha-1 domain of the class I MHC  
 CC HLA-B\*62. These sequences can be used to extend the period of acceptance  
 CC by a recipient of a transplant from an MHC unmatched donor. The peptides  
 CC are administered to a patient in conjunction with a subtherapeutic amount  
 CC of an immunosuppressant. This is administered to the patient for a  
 CC limited period of time (compared to the lifetime administration for  
 CC current treatments). The peptides particularly modulate (or inhibit) the  
 CC activity of the cytotoxic T lymphocytes (CTLs) of the patient.  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLRGY 10  
 |||||  
 DB 16 RESLNRLRGY 25

## RESULT 12

R95431  
 ID R95431 standard; peptide; 25 AA.  
 AC R95431;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B\*62.60-84.  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Compns. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 12: 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B\*62.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLRGY 10  
 |||||  
 DB 16 RESLNRLRGY 25

## RESULT 13

R95419  
 ID R95419 standard; peptide; 25 AA.  
 AC R95419;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B\*62.60-84.  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Compns. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 9: 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B\*62.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLRGY 10  
 |||||  
 DB 16 RESLNRLRGY 25

## RESULT 14

R12464  
 ID R12464 standard; Protein; 362 AA.  
 AC R12464;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-B\*55 antigen.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I.  
 OS Homo sapiens.  
 PN J03112486-A.

PD 14-MAY-1991.  
 PF 22-SEP-1989; 247697.  
 PR 22-SEP-1989; JP-247697.  
 PA (OLIU ) OLYMPUS OPTICAL KK.  
 DR WPI; 91-182991/25.  
 DR N-PSDB; Q12115.  
 PT HLA-B35 gene - used in DNA probe and transformant cells for  
 immunising animals, for developing monoclonal antibody.  
 PS Claim 1; Page 1; lpp; Japanese.  
 CC Probes comprising part of the sequence encoding this sequence can  
 be used to identify Class I genes. The DNA can be expressed for  
 immunisation of animals and prodn. of monoclonal antibodies  
 CC specific for the HLA-B35 antigen. See also J03112485 and J03112487.  
 SQ Sequence 362 AA;

Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.031;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLRNLRGY 10  
 | | | | | | | | | |  
 DB 99 RESLRNLRGY 108

## RESULT 15

R41206  
 ID R41206 standard; peptide; 25 AA.  
 AC R41206;  
 DT 15-MAR-1994 (first entry)  
 DE Peptide fragment of Class I HLA peptide.  
 KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
 OS parasitic disease; cytotoxic T lymphocyte; modulation.  
 OS Synthetic.  
 PN WO9317699-A.  
 PD 16-SEP-1993.  
 PF 25-FEB-1993; U01758.  
 PR 02-MAR-1992; US-844716.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger CA, Krensky AM;  
 DR WPI; 93-303134/38.  
 PT New peptide(s) based on Class I HLA antigen domains - used for  
 modulating cytotoxic T-lymphocyte activity towards targets  
 PS Claim 9; Page 53; 61pp; English.  
 CC The peptide (or a fragment of at least 10 amino acids, joined at at  
 least one terminus to a sequence other than that of wild type HLA  
 antigen) is used to modulate cytotoxic T-lymphocyte (CTL) activity,  
 either by inhibition or stimulation. It can be used for  
 inhibiting CTL toxicity in transplantations, for inducing CTL  
 activity in parasitic diseases and neoplasia and in studies on viral  
 infection. The peptide can also be used for identifying CTLs which  
 bind to it and removing subsets of CTLs from a T-cell composition.  
 SQ Sequence 25 AA;

Query Match 86.3%; Score 44; DB 1; Length 25;  
 Best Local Similarity 90.0%; Pred. No. 0.04;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RESLRNLRGY 10  
 | | | | | | | | | |  
 DB 16 RVSLRNLRGY 25

Search completed: February 8, 2000, 04:05:44  
 Job time: 9361 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 18:04:41 : Search time 111.22 seconds  
(without alignments)  
4.241 Million cell updates/sec

Title: US-08-653-294-34

Perfect score: 51

Sequence: 1 RESLNRLRGY 10

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR\_62:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	137	2	I38875
2	51	100.0	137	2	I38876
3	51	100.0	137	2	I38860
4	51	100.0	137	2	I38874
5	51	100.0	181	2	I59188
6	51	100.0	270	1	HLHU40
7	51	100.0	274	2	I68774
8	51	100.0	274	2	S24439
9	51	100.0	300	2	I68701
10	51	100.0	308	2	I36956
11	51	100.0	350	2	I68747
12	51	100.0	350	2	I54308
13	51	100.0	354	2	S24436
14	51	100.0	354	2	S24437
15	51	100.0	354	2	S24438
16	51	100.0	354	2	S24440
17	51	100.0	354	2	S24433
18	51	100.0	358	2	S03538
19	51	100.0	361	2	I54418
20	51	100.0	362	1	HLHUB7
21	51	100.0	362	2	S77966
22	51	100.0	362	2	G01230
23	51	100.0	362	2	S16789
24	51	100.0	362	2	A45880
25	51	100.0	362	2	I37519
26	51	100.0	362	2	I37520
27	51	100.0	362	2	I37522
28	51	100.0	362	2	I84431
29	51	100.0	362	2	I72755
30	51	100.0	362	2	I84488

31	51	100.0	362	2	I72752
32	51	100.0	362	2	I72753
33	51	100.0	362	2	I72754
34	51	100.0	362	2	I38437
35	51	100.0	362	2	I37492
36	51	100.0	362	2	I54505
37	51	100.0	362	2	I56130
38	51	100.0	362	2	I36962
39	51	100.0	362	2	I37475
40	51	100.0	362	2	I36149
41	51	100.0	362	2	I59655
42	51	100.0	362	2	I59651
43	51	100.0	362	2	I61865
44	51	100.0	362	2	I61859
45	51	100.0	362	2	I61860

## ALIGNMENTS

### RESULT 1

I38875

MHC class I antigen - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I38875

R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stef

Immunogenetics 42, 19-27, 1995

A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.

A:Reference number: I38860; MUID:95317819

A:Accession: I38875

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-137 <RES>

A:Cross-references: EMBL:U15639; NID:930332; PIDN:AAA74046.1; PID:930333

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 100.0%; Score 51; DB 2; Length 137;

Best Local Similarity 100.0%; Pred. No. 0.0045;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLRGY 10

|||||

DB 40 RESLNRLRGY 49

### RESULT 2

I38876

MHC class I antigen - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I38876

R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stef

Immunogenetics 42, 19-27, 1995

A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.

A:Reference number: I38860; MUID:95317819

A:Accession: I38876

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-137 <RES>

A:Cross-references: EMBL:U15640; NID:930334; PIDN:AAA74047.1; PID:930335

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match

Best Local Similarity 100.0%; Score 51; DB 2; Length 137;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLRGY 10

|||||

DB 40 RESLNRLRGY 49

HLA-B*5501 - human
HLA-B*5502 - human
HLA-B*5601 - human
MHC class I histoc
HLA-B alpha-chain
Lymphocyte antigen
HLA-B*5401 - human
MHC class I protei
HLA-B35 variant -
Lymphocyte antigen
Lymphocyte antigen
Lymphocyte antigen
MHC HLA-B*42, HLA-
MHC HLA-B14 chain
MHC HLA-B18 chain

## RESULT 3

I38860  
MHC class I antigen - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
C:Accession: I38860  
R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stefani  
Immunogenetics 42, 19-27, 1995  
A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.  
A:Reference number: I38860; MUID:95317819  
A:Accession: I38860  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-137 <RES>  
A:Cross-references: EMBL:U14756; NID:g930328; PIDN:AA50171.1; PID:g930329  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 100.0%; Score 51; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 0.0045;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLGY 10  
|||||  
Db 40 RESLNRLGY 49

## RESULT 4

I38874  
MHC class I antigen - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
C:Accession: I38874  
R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stefani  
Immunogenetics 42, 19-27, 1995  
A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.  
A:Reference number: I38860; MUID:95317819  
A:Accession: I38874  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-137 <RES>  
A:Cross-references: EMBL:U15638; NID:g930330; PIDN:AAA74045.1; PID:g930331  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 100.0%; Score 51; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 0.0045;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLGY 10  
|||||  
Db 40 RESLNRLGY 49

## RESULT 5

I59188  
MHC cell surface glycoprotein - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999  
C:Accession: I59188  
R:Bronson, S.K.; Pei, J.; Taillon-Miller, P.; Chorney, M.J.; Geraghty, D.E.; Chaplin, D.  
Proc. Natl. Acad. Sci. U.S.A. 88, 1676-1680, 1991  
A:Title: Isolation and characterization of yeast artificial chromosome clones linking the  
A:Reference number: I59188; MUID:91156671  
A:Accession: I59188  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-181 <RES>  
A:Cross-references: GB:M59841; NID:g187697; PIDN:AAA59623.1; PID:g187698  
C:Gene: GDB:HLA-B  
A:Cross-references: GDB:I20048; OMIM:142830

A:Map position: 6p21.3-6p21.3

A:Introns: 90/1  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein

Query Match 100.0%; Score 51; DB 2; Length 181;  
Best Local Similarity 100.0%; Pred. No. 0.0061;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLGY 10  
|||||  
Db 74 RESLNRLGY 83

## RESULT 6

HLH040  
MHC class I histocompatibility antigen HLA-B40 alpha chain - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 02-Sep-1997  
C:Accession: A02186  
R:Lopez de Castro, J.A.; Bragado, R.; Strong, D.M.; Strominger, J.L.  
Biochemistry 22, 3961-3969, 1983  
A:Title: Primary structure of papain-solubilized human histocompatibility antigen HLA  
A:Reference number: A02186; MUID:84000412  
A:Accession: A02186  
A:Molecule type: protein  
A:Residues: 1-270 <LOP>  
C:Comment: Comparison of a number of class I allelic sequences (-B40 vs. -B7 and -A2  
y restricted to the alpha-1 and alpha-2 domains. The most conspicuous clustering of v  
ossible alloantigenic determinants of these antigens.  
C:Genetics:  
A:Gene: GDB:HLA-B  
A:Cross-references: GDB:I20048; OMIM:142830  
C:Map position: 6p21.3-6p21.3  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: duplication; glycoprotein; heterodimer; membrane protein; transplantation  
F:1-90/Domain: alpha-1 <EX1>  
F:91-181/Domain: alpha-2 <EX2>  
F:195-260/Domain: immunoglobulin homology <IMM>  
F:86/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:101-163,202-258/disulfide bonds: #status predicted

Query Match 100.0%; Score 51; DB 1; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.0093;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLGY 10  
|||||  
Db 75 RESLNRLGY 84

## RESULT 7

I68774  
MHC HLA-B39 chain - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
C:Accession: I68774  
R:Mueller, C.A.; Engler-Blum, G.; Gekeler, V.; Steiert, I.; Weiss, E.; Schmidt, H.  
Immunogenetics 30, 200-207, 1989  
A:Title: Genetic and serological heterogeneity of the supertypic HLA-B locus specific  
A:Reference number: I54463; MUID:89379286  
A:Accession: I68774  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-274 <RES>  
A:Cross-references: GB:M29865; NID:g187676; PIDN:AAA36223.1; PID:g187677  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 100.0%; Score 51; DB 2; Length 274;  
Best Local Similarity 100.0%; Pred. No. 0.0095;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10  
|||||

Db 75 RESLNRLGY 84

## RESULT 8

S24439  
Class I histocompatibility antigen HLA-B-3901 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence\_revision 01-Nov-1996 #text\_change 07-Nov-1997

C:Accession: S24439

R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Garbe

Nature 357, 329-333, 1992

A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indicate

A:Reference number: S24027; MUID:92269956

A:Accession: S24439

A:Molecule type: mRNA

A:Residues: 1-274 <WAT>

C:Genetics:

A:Gene: HLA-B-3901

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: glycoprotein

F:1-274/Product: class I histocompatibility antigen HLA-B-3901 #status predicted <MAT>

F:1-90/Domain: alpha-1 #status predicted <EX1>

F:91-182/Domain: alpha-2 #status predicted <EX2>

F:186-261/Domain: immunoglobulin homology <IMM>

F:86/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:101-164,203-259/Disulfide bonds: #status predicted

## Query Match

Best Local Similarity 100.0%; Score 51; DB 2; Length 274;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10  
|||||

Db 75 RESLNRLGY 84

## RESULT 9

I68701

cell surface antigen - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I68701

R:Arnot, D.; Lillie, J.W.; Auffray, C.; Kappes, D.; Strominger, J.L.

Immunogenetics 20, 237-252, 1984

A:Title: Inter-locus and intra-allelic polymorphisms of HLA class I antigen gene mRNA.

A:Reference number: I5412; MUID:84287690

A:Accession: I68701

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-300 <RES>

A:Cross-references: GB:M27540; NID:q187733; PIDN:AAA59638.1; PID:q386890

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: surface antigen

## Query Match

Best Local Similarity 100.0%; Score 51; DB 2; Length 300;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10  
|||||

Db 37 RESLNRLGY 46

## RESULT 10

I36956

MHC class II chain - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999

C:Accession: I36956

R:Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.

J. Immunol. 142, 3937-3950, 1989

A:Title: Diversity and diversification of HLA-A,B,C alleles.

A:Reference number: I36956; MUID:89235215

A:Accession: I36956

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-308 <RES>

A:Cross-references: GB:M24044; NID:q176812; PIDN:AAA35423.1; PID:q176813

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

## Query Match

Best Local Similarity 100.0%; Score 51; DB 2; Length 308;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10  
|||||

Db 45 RESLNRLGY 54

## RESULT 11

I68747

MHC class I lymphocyte antigen - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I68747

R:Polhla, H.; Kuon, W.; Tabaczewski, P.; Doerner, C.; Weiss, E.H.

Immunogenetics 29, 297-307, 1989

A:Title: Allelic variation in HLA-B and HLA-C sequences and the evolution of the HLA-

A:Reference number: I54457; MUID:89233295

A:Accession: I68747

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-350 <RES>

A:Cross-references: GB:M28204; NID:q576472; PIDN:AAA53257.1; PID:q576473

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

## Query Match

Best Local Similarity 100.0%; Score 51; DB 2; Length 350;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10  
|||||

Db 87 RESLNRLGY 96

## RESULT 12

I54308

MHC class I - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999

C:Accession: I54308

R:Rodriguez, S.G.; Johnson, A.H.; Hurley, C.K.

Hum. Immunol. 37, 192-194, 1993

A:Title: Molecular characterization of HLA-B\*71 from an African American individual.

A:Reference number: I54308; MUID:94064392

A:Accession: I54308

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-350 <RES>

A:Cross-references: GB:I07950; NID:q307236; PIDN:AAA59683.1; PID:q307237

C:Genetics:

A:Gene: GDB:HLA-B

A:Cross-references: GDB:120048; OMIM:142830

A:Map position: 6p21.3-6p21.3

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

## Query Match

Best Local Similarity 100.0%; Score 51; DB 2; Length 350;

```
Best Local Similarity 100.0%; Score 51; DB 2; Length 354;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10
   |||||
Db 99 RESLNRLGY 108

RESULT 13
S24436
class I histocompatibility antigen HLA-B-3504 precursor - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1993 #sequence_revision 18-Jul-1997 #text_change 07-Nov-1997
C:Accession: S24436
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Garbe
Nature 357, 329-333, 1992
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indicate
A:Reference number: S24027; MUID:92269956
A:Accession: S24436
A:Molecule type: mRNA
A:Residues: 1-354 <MAT>
C:Genetics:
A:Gene: HLA-B-3504
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: glycoprotein; transmembrane protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:17-354/Product: class I histocompatibility antigen HLA-B-3504 #status predicted <MAT>
F:17-299/Domain: extracellular #status predicted <EXT>
F:107-198/Domain: alpha-2 #status predicted <EX2>
F:212-277/Domain: immunoglobulin homology <IMM>
F:300-323/Domain: transmembrane #status predicted <TM>
F:324-354/Domain: intracellular #status predicted <INT>
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:117-180,219-275/Disulfide bonds: #status predicted

Query Match 100.0%; Score 51; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10
   |||||
Db 91 RESLNRLGY 100

RESULT 14
S24437
class I histocompatibility antigen HLA-B-4802 precursor - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1993 #sequence_revision 18-Jul-1997 #text_change 07-Nov-1997
C:Accession: S24437
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Garbe
Nature 357, 329-333, 1992
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indicate
A:Reference number: S24027; MUID:92269956
A:Accession: S24437
A:Molecule type: mRNA
A:Residues: 1-354 <MAT>
C:Genetics:
A:Gene: HLA-B-4802
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: glycoprotein; transmembrane protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:17-354/Product: class I histocompatibility antigen HLA-B-4802 #status predicted <MAT>
F:17-299/Domain: extracellular #status predicted <EXT>
F:107-198/Domain: alpha-2 #status predicted <EX2>
F:212-277/Domain: immunoglobulin homology <IMM>
F:300-323/Domain: transmembrane #status predicted <TM>
F:324-354/Domain: intracellular #status predicted <INT>
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:117-180,219-275/Disulfide bonds: #status predicted
```

```
Query Match 100.0%; Score 51; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10
   |||||
Db 91 RESLNRLGY 100

RESULT 15
S24438
class I histocompatibility antigen HLA-B-4801 precursor - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1993 #sequence_revision 18-Jul-1997 #text_change 07-Nov-1997
C:Accession: S24438
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Ga
Nature 357, 329-333, 1992
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indic
A:Reference number: S24027; MUID:92269956
A:Accession: S24438
A:Molecule type: mRNA
A:Residues: 1-354 <MAT>
C:Genetics:
A:Gene: HLA-B-4801
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: glycoprotein; transmembrane protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:17-354/Product: class I histocompatibility antigen HLA-B-4801 #status predicted <MA
F:17-299/Domain: extracellular #status predicted <EXT>
F:107-198/Domain: alpha-2 #status predicted <EX2>
F:212-277/Domain: immunoglobulin homology <IMM>
F:300-323/Domain: transmembrane #status predicted <TM>
F:324-354/Domain: intracellular #status predicted <INT>
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:117-180,219-275/Disulfide bonds: #status predicted

Query Match 100.0%; Score 51; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNRLGY 10
   |||||
Db 91 RESLNRLGY 100

Search completed: February 7, 2000, 18:04:41
Job time: 22207 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:26:01 ; Search time 68.63 seconds  
(without alignments)  
4.352 Million cell updates/sec

Title: US-08-653-294-34

Perfect score: 51

Sequence: 1 RESLNRLGY 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	51	100.0	270	1	1B33_HUMAN	P01890 homo sapien
2	51	100.0	354	1	1B24_HUMAN	P30470 homo sapien
3	51	100.0	362	1	1A02_PANTR	P16210 pan troglod
4	51	100.0	362	1	1B02_HUMAN	P01889 homo sapien
5	51	100.0	362	1	1B02_PANTR	P13751 pan troglod
6	51	100.0	362	1	1B04_HUMAN	P30460 homo sapien
7	51	100.0	362	1	1B07_HUMAN	P30462 homo sapien
8	51	100.0	362	1	1B08_HUMAN	P30463 homo sapien
9	51	100.0	362	1	1B10_HUMAN	P30464 homo sapien
10	51	100.0	362	1	1B11_HUMAN	P30465 homo sapien
11	51	100.0	362	1	1B12_HUMAN	P30513 homo sapien
12	51	100.0	362	1	1B13_HUMAN	P30466 homo sapien
13	51	100.0	362	1	1B21_HUMAN	P30685 homo sapien
14	51	100.0	362	1	1B22_HUMAN	P30468 homo sapien
15	51	100.0	362	1	1B23_HUMAN	P30469 homo sapien
16	51	100.0	362	1	1B25_HUMAN	P30471 homo sapien
17	51	100.0	362	1	1B26_HUMAN	P30472 homo sapien
18	51	100.0	362	1	1B27_HUMAN	P30473 homo sapien
19	51	100.0	362	1	1B28_HUMAN	P30474 homo sapien
20	51	100.0	362	1	1B31_HUMAN	P30475 homo sapien
21	51	100.0	362	1	1B32_HUMAN	P30476 homo sapien
22	51	100.0	362	1	1B34_HUMAN	Q04826 homo sapien
23	51	100.0	362	1	1B35_HUMAN	P30477 homo sapien
24	51	100.0	362	1	1B36_HUMAN	P30478 homo sapien
25	51	100.0	362	1	1B38_HUMAN	P30479 homo sapien
26	51	100.0	362	1	1B39_HUMAN	P30480 homo sapien
27	51	100.0	362	1	1B43_HUMAN	P30483 homo sapien
28	51	100.0	362	1	1B46_HUMAN	P30486 homo sapien
29	51	100.0	362	1	1B48_HUMAN	P30488 homo sapien
30	51	100.0	362	1	1B55_HUMAN	P30492 homo sapien
31	51	100.0	362	1	1B56_HUMAN	P30493 homo sapien
32	51	100.0	362	1	1B57_HUMAN	P30494 homo sapien
33	51	100.0	362	1	1B58_HUMAN	P30495 homo sapien
34	51	100.0	362	1	1B59_HUMAN	P30496 homo sapien

35	51	100.0	362	1	1B63_HUMAN	P30498 homo sapien
36	44	86.3	342	1	1CXX_HUMAN	P10321 homo sapien
37	44	86.3	362	1	1B44_HUMAN	P30484 homo sapien
38	44	86.3	366	1	1C01_HUMAN	P30499 homo sapien
39	44	86.3	366	1	1C01_PANTR	P30686 pan troglod
40	44	86.3	366	1	1C02_HUMAN	P30500 homo sapien
41	44	86.3	366	1	1C05_HUMAN	P04222 homo sapien
42	44	86.3	366	1	1C06_HUMAN	P30503 homo sapien
43	44	86.3	366	1	1C11_HUMAN	Q29631 homo sapien
44	44	86.3	366	1	1C12_HUMAN	P30505 homo sapien
45	44	86.3	366	1	1C13_HUMAN	P30506 homo sapien

## ALIGNMENTS

RESULT 1  
ID 1B33\_HUMAN STANDARD; PRT; 270 AA.  
AC P01890;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW\*60(B\*40) B\*4001 ALPHA CHAIN (FRAGMENT).  
DE (FRAGMENT).  
GN HLA-B OR HLAB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE.  
RX MEDLINE: 84000412.  
RA LOPEZ DE CASTRO J.A., BRAGADO R., SIROG D.M., SIROMINGER J.L.;  
RT "Primary structure of papain-solubilized human histocompatibility antigen HLA-B\*40 (-BW60). An outline of alloantigenic determinants.";  
RL Biochemistry 22:3961-3969(1983).  
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).  
CC PIR: A02186; HLH040.  
DR HSSP: P30460; LABB.  
DR MIN: 142830; -  
DR PROSITE: PS00290; IG\_MHC; 1.  
DR PFAM: PF00047; Ig; 1.  
DR PFAM: PF00129; MHC\_1; 1.  
DR MHC I; Glycoprotein.  
KW DOMAIN 1 90  
FT DOMAIN 91 181  
FT DOMAIN 182 >270  
FT CARBOHYD 86 86  
FT DISULFID 101 163  
FT DISULFID 202 258  
FT NON\_TER 270 270  
SQ SEQUENCE 270 AA; BFE44EFF CRC32;

Query Match 100.0%; Score 51; DB 1; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.0048;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RESLNRLGY 10  
Db 75 RESLNRLGY 84

RESULT 2  
ID 1B24\_HUMAN STANDARD; PRT; 354 AA.  
AC P30470;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3504 ALPHA CHAIN  
DE PRECURSOR (FRAGMENT).  
GN HLA-B OR HLAB.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homidae; Homo.

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92269956.

RA WATKINS D.I., MCADAM S.N., LIU X., STANG C.R., MILFORD E.L.,  
RA LEVINE C.G., GUBES A.L., DOGON A.L., LORD C.I., GHIM S.H.,  
RA TROUP G.M., HUGHES A.L., LETVIN N.L.;

RT "New recombinant HLA-B alleles in a tribe of South American  
RL Amerindians indicate rapid evolution of MHC class I loci";  
RL Nature 357:329-333(1992).

CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M86403; ; NOT\_ANNOTATED\_CDS.  
CC HSP; P30685; 1A1N.  
CC MTM; 142830; ;  
CC PROSITE; PS00290; IG\_MHC; 1.  
CC PFAM; PF00047; 1g; 1.  
CC PFAM; PF00129; MHC\_I; 1.  
CC MHC I; Transmembrane; Glycoprotein; Signal.  
CC NON\_TER 1 1  
CC SIGNAL <1 16  
CC CHAIN 17 354

FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
FT B-35 B\*3504 ALPHA CHAIN.  
FT EXTRACELLULAR ALPHA-1.  
FT EXTRACELLULAR ALPHA-2.  
FT EXTRACELLULAR ALPHA-3.  
FT CONNECTING PEPTIDE.  
FT TRANSMEM 301 324  
FT CYTOPLASMIC TAIL.  
FT DOMAIN 325 354  
FT CARBOHYD 102 102  
FT DISULFID 117 180  
FT DISULFID 219 275  
FT SEQUENCE 354 AA; 6564795A CRC32;

Query Match 100.0%; Score 51; DB 1; Length 354;  
Best Local Similarity 100.0%; Pred. No. 0.0065;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
| | | | | | | | | |  
Db 91 RESLNLRGY 100

RESULT 3  
1A02\_PANTR STANDARD; PRT; 362 AA.  
AC P16210;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-5 ALPHA CHAIN PRECURSOR.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homidae; Pan.  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE; 90201944.  
RA LAWOR D.A., WARREN E., WARD F.E., PARHAM P.;  
RT "Comparison of class I MHC alleles in humans and apes";  
RL Immunol. Rev. 113:147-185(1990).

CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M30679; AAA87971.1; ;  
CC HSP; P30685; 1A1N.  
CC PROSITE; PS00290; IG\_MHC; 1.  
CC PFAM; PF00047; 1g; 1.  
CC PFAM; PF00129; MHC\_I; 1.  
CC MHC I; Transmembrane; Glycoprotein; Signal.  
CC SIGNAL 1 24  
CC CHAIN 25 362

FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
FT A-5 ALPHA CHAIN.  
FT EXTRACELLULAR ALPHA-1.  
FT EXTRACELLULAR ALPHA-2.  
FT EXTRACELLULAR ALPHA-3.  
FT CONNECTING PEPTIDE.  
FT TRANSMEM 309 332  
FT CYTOPLASMIC TAIL.  
FT DOMAIN 333 362  
FT CARBOHYD 125 188  
FT DISULFID 227 283  
FT DISULFID 110 110  
FT SEQUENCE 362 AA; 40487 MW; 9756CE8A CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 0.0066;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
| | | | | | | | | |  
Db 99 RESLNLRGY 108

RESULT 4  
1B02\_HUMAN STANDARD; PRT; 362 AA.  
ID 1B02\_HUMAN  
AC P01889;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-7 B\*0702 ALPHA CHAIN  
DE PRECURSOR (B7.2).  
GN HLA-B OR HLAB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90207291.  
RA ENNIS P.D., ZEMMOUR J., SALTER R.D., PARHAM P.;  
RT "Rapid cloning of HLA-A,B CDNA by using the polymerase chain  
RL reaction: frequency and nature of errors produced in amplification";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:2833-2837(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90315860.  
RA PARHAM P., BENJAMIN R.J., CHEN B.P., CLAYBERGER C., ENNIS P.D.,  
RA KRENSKY A.M., LAWOR D.A., LITTMAN D.R., NORMENT A.M., ORR H.T.,  
RA SALTER R.D., ZEMMOUR J.

RT "Diversity of class I HLA molecules: functional and evolutionary  
 interactions with T cells.";  
 RL Cold Spring Harb. Symp. Quant. Biol. 54:529-543(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85287366.  
 RA SOOD A.K., PAN J., BIRO P.A., PEREIRA D., SRIVASTAVA R., REDDY V.B.,  
 RA DUCMAN B.W., WEISSMAN S.M.;  
 RT "Structure and polymorphism of class I MHC antigen mRNA.";  
 RL Immunogenetics 22:101-121(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA ELLEXSON M.E., ZHANG L., HILDEBRAND W.H.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 25-295.  
 RX MEDLINE; 80088278.  
 RA ORR H.T., LOPEZ DE CASTRO J.A., LANCET D., STROMINGER J.L.;  
 RT "Complete amino acid sequence of a papain-solubilized human  
 RT histocompatibility antigen, HLA-B\*7.2. Sequence determination and  
 search for homologies.";  
 RL Biochemistry 18:5711-5720(1979).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M32317; AAA36230.1; -;  
 DR EMBL; M16102; AAA59622.1; ALT\_SEQ.  
 DR EMBL; U29057; AAA91229.1; -;  
 DR PIR; A02185; HLHUB7.  
 DR PIR; B35997; B35997.  
 DR HSSP; P30460; LAGB.  
 DR MIM; 142830; -;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT B-7 B\*0702 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 309 CONNECTING PEPTIDE.  
 FT TRANSMEM 310 333 CYTOPLASMIC TAIL.  
 FT DOMAIN 334 362  
 FT CARBOHYD 110 110  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 FT CONFLICT 15 18 AALA -> GPW (IN REF. 3).  
 FT CONFLICT 266 266 Q -> E (IN REF. 5).  
 FT CONFLICT 268 268 W -> S (IN REF. 3).  
 FT CONFLICT 297 297 R -> G (IN REF. 3).  
 FT CONFLICT 314 315 GL -> RP (IN REF. 3).  
 SQ SEQUENCE 362 AA; 40460 MW; 87B2ED84 CRC32;  
 Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RESLRNLRGY 10  
 Db 99 RESLRNLRGY 108  
 RESULT 6  
 ID 1B04\_HUMAN  
 AC P30460; STANDARD; PRT; 362 AA.

RESULT 5  
 ID 1B02\_PANTR  
 ID 1B02\_PANTR STANDARD; PRT; 362 AA.  
 AC P13751;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-2 ALPHA CHAIN PRECURSOR.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89030641.  
 RA MAYER W.E., JONKER M., KLEIN D., IVANYI P., VAN SEVENTER G.,  
 RA KLEIN J.;  
 RT "Nucleotide sequences of chimpanzee MHC class I alleles: evidence for  
 RT trans-species mode of evolution.";  
 RL EMBO J. 7:2765-2774(1988).  
 RN [2]  
 RP REVISIONS.  
 RA MAYER W.;  
 RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13116; CAA31508.1; -;  
 DR PIR; S03538; S03538.  
 DR HSSP; P30685; IAIN.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT B-2 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 308 CONNECTING PEPTIDE.  
 FT TRANSMEM 309 332 CYTOPLASMIC TAIL.  
 FT DOMAIN 333 362  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40488 MW; 4BF65A6C CRC32;  
 Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RESLRNLRGY 10  
 Db 99 RESLRNLRGY 108  
 RESULT 6  
 ID 1B04\_HUMAN  
 AC P30460; STANDARD; PRT; 362 AA.





OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 8935215.  
 RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
 RA "Diversity and diversification of HLA-A,B,C alleles";  
 RT J. Immunol. 142:3937-3950(1989).  
 RL -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M24032; AAS59664.1; -.  
 DR HSP; P30460; IAGB.  
 DR MIM; 142830; -.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT BW-65(B-14) B\*1402 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 309 CONNECTING PEPTIDE.  
 FT TRANSMEM 310 333  
 FT DOMAIN 334 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; BE68AC9E CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLRNLRGY 10  
 |||||  
 Db 99 RESLRNLRGY 108

RESULT 9  
 ID IB10\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30464;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-75(B-15) B\*1502 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92196792.  
 RA "The HLA-B\*75 subtype of B15: molecular characterization and  
 RT comparison with crossreacting antigens";  
 RL Tissue Antigens 38:186-190(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC TISSUE-BLOOD;  
 RX MEDLINE; 96359309.  
 RA LIN L., TORUNAGA K., TANAKA H., NAKAJIMA F., IMANISHI T.,  
 RA KASHIWASE K., BANNAI M., MIZUNO S., AKAZA T., TADOKORO K.,  
 RA SHIBATA Y., JUJI T.;  
 RT "Further molecular diversity in the HLA-B15 group";  
 RL Tissue Antigens 47:265-274(1996).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M75138; AAS59630.1; -.  
 DR EMBL; D50293; BAA08824.1; -.  
 DR HSP; P30685; IALN.  
 DR MIM; 142830; -.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT BW-75(B-15) B\*1502 ALPHA CHAIN.  
 FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 299 309 CONNECTING PEPTIDE.  
 FT TRANSMEM 310 333  
 FT DOMAIN 334 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 8CF9BCD0 CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLRNLRGY 10  
 |||||  
 Db 99 RESLRNLRGY 108

RESULT 10  
 ID IB11\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30465;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-72(BW-70) B\*1503 ALPHA  
 DE CHAIN PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93056508.  
 RA MADRIGAL J.A., BELICH M.P., HILDEBRAND W.H., BENJAMIN R.J.,  
 RA LITTLE A.-M., ZEMMOUR J., ENNIS P.D., WARD F.E., PETZL-ERLER M.L.,  
 RA MARTELL R.W., DU TOIT E.D., PARHAM P.;  
 RT "Distinctive HLA-A,B antigens of black populations formed by  
 RT interallelic conversion";

```

RL J. Immunol. 149:3411-3415(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61709; CAA43878.1; -
CC PIR; S16789; S16789.
CC HSP; P30685; IAIN.
CC MIM; 142830.
CC PROSITE; PS00290; IG_MHC; 1.
CC PFAM; PF00047; I; 1.
CC PFAM; PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC BW-72(BW-70) B*1503 ALPHA CHAIN.
CC DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
CC DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
CC DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
CC DOMAIN 299 309 CONNECTING PEPTIDE.
CC TRANSMEM 310 333
CC DOMAIN 334 362 CYTOPLASMIC TAIL.
CC CARBOHYD 110 110 BY SIMILARITY.
CC DISULFID 125 188 BY SIMILARITY.
CC DISULFID 227 283 BY SIMILARITY.
CC SEQUENCE 362 AA; 40387 MW; 99D70546 CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLRNLRGY 10
DB 99 RESLRNLRGY 108

RESULT 11
1B12_HUMAN
ID 1B12_HUMAN STANDARD; PRT; 362 AA.
AC P30513.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-62 B*1504 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RX MEDLINE; 92269956.
RA WATKINS D.I., MCADAM S.N., LIU X., STANG C.R., MILFORD E.L.,
RA LEVINE C.G., GABER T.L., DOGAN A.L., LORD C.I., GHIM S.H.,
RA TROUP G.M., HUGHES A.L., LETVIN N.L.;
RT "New recombinant HLA-B alleles in a tribe of South American
RT Amerindians indicate rapid evolution of MHC class I loci.";
RL Nature 357:329-333(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RA RAMOS M., BARBER D.F., LAYRISSE Z., DE CASTRO J.A.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

```

```

CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84382; AAA59632.1; -
CC PIR; S24433; S24433.
CC HSP; P30685; IAIN.
CC MIM; 142830.
CC PROSITE; PS00290; IG_MHC; 1.
CC PFAM; PF00047; I; 1.
CC PFAM; PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC B-62 B*1504 ALPHA CHAIN.
CC DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
CC DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
CC DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
CC DOMAIN 299 309 CONNECTING PEPTIDE.
CC TRANSMEM 310 333
CC DOMAIN 334 362 CYTOPLASMIC TAIL.
CC CARBOHYD 110 110 BY SIMILARITY.
CC DISULFID 125 188 BY SIMILARITY.
CC DISULFID 227 283 BY SIMILARITY.
CC SEQUENCE 362 AA; 40406 MW; E80FC24C CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLRNLRGY 10
DB 99 RESLRNLRGY 108

RESULT 12
1B13_HUMAN
ID 1B13_HUMAN STANDARD; PRT; 362 AA.
AC P30466.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-18 B*1801 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RX MEDLINE; 89235215.
RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M24039; AAA59662.1; -  
 DR HSP; P30460; IAGB.  
 DR MIM; 142830;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT  
 FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT B-18 B\*1801 ALPHA CHAIN.  
 FT EXTRACELLULAR ALPHA-1.  
 FT EXTRACELLULAR ALPHA-2.  
 FT EXTRACELLULAR ALPHA-3.  
 FT CONNECTING PEPTIDE.  
 FT  
 FT CYTOPLASMIC TAIL.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 SQ SEQUENCE 362 AA; 40275 MW; 038EC3FC CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RESLNRLRGY 10  
 DB 99 RESLNRLRGY 108  
 |||||

RESULT 13  
 1B21\_HUMAN STANDARD; PRT; 362 AA.  
 ID 1B21\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30685;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3501 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 89339610.  
 RA OBA T., HAYASHI H., KARAKI S., TANABE M., KANO K., TAKIGUCHI M.;  
 RT "The structure of HLA-B\*35 suggests that it is derived from HLA-B\*58  
 RT by two genetic mechanisms.";  
 RL Immunogenetics 30:76-80(1989).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 25-300.  
 RX MEDLINE; 96209671.  
 RA SMITH K.J., REID S.W., STUART D.I., MCMICHAEL A.J., JONES E.Y.,  
 RA BELL J.I.;  
 RT "An altered position of the alpha 2 helix of MHC class I is revealed  
 RT by the crystal structure of HLA-B\*3501.";  
 RL Immunity 4:203-214(1996).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).  
 RA MENSSSEN R., ORTH P., ZIEGLER A., SAENGER W.;  
 RL Submitted (APR-1998) to the PDB data bank.  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M28115; AAA59617.1;  
 DR EMBL; M28109; AAA59617.1; JOINED.  
 DR EMBL; M28110; AAA59617.1; JOINED.  
 DR EMBL; M28111; AAA59617.1; JOINED.  
 DR EMBL; M28112; AAA59617.1; JOINED.  
 DR EMBL; M28113; AAA59617.1; JOINED.  
 DR EMBL; M28114; AAA59617.1; JOINED.  
 DR PIR; A45880; A45880.  
 DR PDB; IAIN; 08-APR-98.  
 DR PDB; IAGB; 18-NOV-98.  
 DR PDB; IAGB; 18-NOV-98.  
 DR MIM; 142830;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT B-35 B\*3501 ALPHA CHAIN.  
 FT EXTRACELLULAR ALPHA-1.  
 FT EXTRACELLULAR ALPHA-2.  
 FT EXTRACELLULAR ALPHA-3.  
 FT CONNECTING PEPTIDE.  
 FT  
 FT CYTOPLASMIC TAIL.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 SQ SEQUENCE 362 AA; 40455 MW; AEGIC675 CRC32;

Query Match 100.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RESLNRLRGY 10  
 DB 99 RESLNRLRGY 108  
 |||||

RESULT 14  
 1B22\_HUMAN STANDARD; PRT; 362 AA.  
 ID 1B22\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30486;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3502 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 91365651.  
 RA CHERTKOFF L.P., HERRERA M., FAINBOIM L., SATZ M.L.;  
 RT "Complete nucleotide sequence of a genomic clone encoding HLA-B\*35  
 RT isolated from a Caucasian individual of Hispanic origin.  
 RT Identification of a new variant of HLA-B\*35.";  
 RL Hum. Immunol. 31:153-158(1991).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).

```

CC -----SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63454; AAA59682.1; -
CC HSSP; P30685; IAIN.
CC MIM; 142830; -
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; ig; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 362
CC FT DOMAIN 25 114
CC FT DOMAIN 115 206
CC FT DOMAIN 207 298
CC FT DOMAIN 299 308
CC FT TRANSMEM 309 332
CC FT DOMAIN 333 362
CC FT CARBOHYD 110 110
CC FT DISULFID 125 188
CC FT DISULFID 227 283
CC FT SEQUENCE 362 AA; 40564 MW; D7B5C2C9 CRC32;
CC
CC Query Match 100.0%; Score 51; DB 1; Length 362;
CC Best Local Similarity 100.0%; Pred. No. 0.0066;
CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 RESLRNLRGY 10
CC DB 99 RESLRNLRGY 108
CC
CC RESULT 15
CC ID 1B23 HUMAN STANDARD; PRT; 362 AA.
CC AC P30469;
CC DT 01-APR-1993 (Rel. 25, Created)
CC DT 01-APR-1993 (Rel. 25, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B*3503 ALPHA CHAIN
CC DE PRECURSOR.
CC GN HLA-B OR HLAB.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP MEDLINE; 92176661.
CC RA ZEMMOUR J., LITTLE A.M., SCHENDEL D.J., PARHAM P.;
CC RT "The HLA-A,B 'negative' mutant cell line CLR expresses a novel
CC RT HLA-B*35 allele, which also has a point mutation in the translation
CC RT initiation codon."
CC RN J. Immunol. 148:1941-1948(1992).
CC [2]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE-BLOOD;
CC RX MEDLINE; 95279930.
CC RA BECK Y., SATZ L., TAKAMIYA Y., NAKAYAMA S., LING L., ISHIKAWA Y.,
CC RA NAGAO T., UCHIDA H., TOKUNAGA K., MULLER C., JUJI T., TAKIGUCHI M.;
CC RT "Polymorphism of human minor histocompatibility antigens: T cell
CC RT recognition of human minor histocompatibility peptides presented by
CC RT HLA-B*35 subtype molecules."
CC RL J. Exp. Med. 181:2037-2048(1995).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.

```

```

CC -----SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81798; AAA59684.1; -
CC HSSP; P30685; IAIN.
CC MIM; 142830; -
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; ig; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 362
CC FT DOMAIN 25 114
CC FT DOMAIN 115 206
CC FT DOMAIN 207 298
CC FT DOMAIN 299 308
CC FT TRANSMEM 309 332
CC FT DOMAIN 333 362
CC FT CARBOHYD 110 110
CC FT DISULFID 125 188
CC FT DISULFID 227 283
CC FT SEQUENCE 362 AA; 40515 MW; 4222D30A CRC32;
CC
CC Query Match 100.0%; Score 51; DB 1; Length 362;
CC Best Local Similarity 100.0%; Pred. No. 0.0066;
CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 RESLRNLRGY 10
CC DB 99 RESLRNLRGY 108
CC
CC Search completed: February 8, 2000, 01:26:01
CC Job time: 1561 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 19:16:17 ; Search time 176.54 Seconds  
(without alignments)  
3.927 Million cell updates/sec

Title: US-08-653-294-34  
Perfect score: 51  
Sequence: 1 RESLNLRGY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL12:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	100.0	81	7	O19523 homo sapien
2	51	100.0	81	7	O19525 homo sapien
3	51	100.0	81	7	O19527 homo sapien
4	51	100.0	81	7	O19529 homo sapien
5	51	100.0	81	7	O19531 homo sapien
6	51	100.0	81	7	O19533 homo sapien
7	51	100.0	81	7	O19535 homo sapien
8	51	100.0	81	7	O19537 homo sapien
9	51	100.0	83	7	O19539 homo sapien
10	51	100.0	83	7	O19541 homo sapien
11	51	100.0	83	7	O19543 homo sapien
12	51	100.0	89	4	O19545 homo sapien
13	51	100.0	89	7	P79615 homo sapien
14	51	100.0	89	7	P79620 homo sapien
15	51	100.0	89	7	O19548 homo sapien
16	51	100.0	89	7	O19549 homo sapien
17	51	100.0	89	7	O19550 homo sapien
18	51	100.0	89	7	O19551 homo sapien
19	51	100.0	89	7	O19640 homo sapien
20	51	100.0	89	7	O77967 homo sapien

21	51	100.0	89	7	O78174 homo sapien
22	51	100.0	89	7	O19638 homo sapien
23	51	100.0	89	7	P79487 homo sapien
24	51	100.0	89	7	O19567 homo sapien
25	51	100.0	89	7	O19576 homo sapien
26	51	100.0	89	7	O19577 homo sapien
27	51	100.0	89	7	O19581 homo sapien
28	51	100.0	89	7	O19583 homo sapien
29	51	100.0	112	7	O92671 homo sapien
30	51	100.0	134	7	O28677 homo sapien
31	51	100.0	137	7	O29660 homo sapien
32	51	100.0	137	7	O29663 homo sapien
33	51	100.0	137	7	O29665 homo sapien
34	51	100.0	137	7	O29664 homo sapien
35	51	100.0	147	7	O29707 homo sapien
36	51	100.0	165	7	P79665 homo sapien
37	51	100.0	165	7	P79666 homo sapien
38	51	100.0	166	7	O19777 homo sapien
39	51	100.0	166	7	O19778 homo sapien
40	51	100.0	172	7	O19776 homo sapien
41	51	100.0	176	7	O29662 homo sapien
42	51	100.0	179	7	O28676 homo sapien
43	51	100.0	180	7	O29678 homo sapien
44	51	100.0	180	7	O77960 homo sapien
45	51	100.0	181	7	O29716 homo sapien

## ALIGNMENTS

RESULT 1  
O19523 PRELIMINARY; PRT; 81 AA.  
AC O19523;  
DT O1-JAN-1998 (TREMBLrel. 05, Created)  
DT O1-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT O1-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAVONGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014769; AAB67807.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
FT NON\_TER 81 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;  
  
Query Match 100.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.013; 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RESLNLRGY 10  
Db 66 RESLNLRGY 75  
|||||||  
  
RESULT 2  
O19525 PRELIMINARY; PRT; 81 AA.  
ID O19525;  
AC O19525;  
DT O1-JAN-1998 (TREMBLrel. 05, Created)  
DT O1-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT O1-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 [1]  
 RN PFAM; PF00129; MHC\_I; 1.  
 RP SEQUENCE FROM N.A.  
 RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF014771; AAB67809.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 81  
 SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 0.013;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 66 RESLNLRGY 75

RESULT 3  
 O19527 PRELIMINARY; PRT; 81 AA.  
 ID O19527  
 AC O19527  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF014773; AAB67811.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 81  
 SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 0.013;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 66 RESLNLRGY 75

RESULT 4  
 O19529 PRELIMINARY; PRT; 81 AA.  
 ID O19529  
 AC O19529  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF014775; AAB67813.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 81  
 SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 0.013;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 66 RESLNLRGY 75

RESULT 5  
 O19531 PRELIMINARY; PRT; 81 AA.  
 ID O19531  
 AC O19531  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF014777; AAB67815.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 81  
 SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 0.013;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
 DB 66 RESLNLRGY 75

RESULT 6  
 O19533 PRELIMINARY; PRT; 81 AA.  
 ID O19533  
 AC O19533  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
 GN HLA-B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
 RA RUNGROUNG E., BEJCHANDRA S.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF014779; AAB67817.1; -  
 DR PFAM; PF00129; MHC\_I; 1.

KW MHC.  
FT NON\_TER. 1 1  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
DB 66 RESLNLRGY 75

RESULT 7  
ID O19535 PRELIMINARY; PRT; 81 AA.  
AC O19535;  
DT 01-JAN-1998 (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGRONG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014781; AAB67819.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
DB 66 RESLNLRGY 75

RESULT 8  
ID O19537 PRELIMINARY; PRT; 81 AA.  
AC O19537;  
DT 01-JAN-1998 (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGRONG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014783; AAB67821.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 100.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
DB 66 RESLNLRGY 75

RESULT 9  
ID O19539 PRELIMINARY; PRT; 83 AA.  
AC O19539;  
DT 01-JAN-1998 (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGRONG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014785; AAB67823.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
SQ SEQUENCE 83 AA; 9731 MW; 24B8D666 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 83;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10  
DB 68 RESLNLRGY 77

RESULT 10  
ID O19541 PRELIMINARY; PRT; 83 AA.  
AC O19541;  
DT 01-JAN-1998 (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGRONG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014787; AAB67825.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
SQ SEQUENCE 83 AA; 9731 MW; 24B8D666 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 83;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 RESLNLRGY 10
DB 68 RESLNLRGY 77

RESULT 11
O19543
ID O19543 PRELIMINARY; PRT; 83 AA.
AC O19543
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,
RA RUNGRONG E., BEJCHANDRA S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF04789; AAB7827.1; -.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 83
SQ SEQUENCE 83 AA; 9731 MW; 24BD6666 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 83;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10
DB 68 RESLNLRGY 77

RESULT 12
Q9Y452
ID Q9Y452 PRELIMINARY; PRT; 89 AA.
AC Q9Y452
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HUMAN LEUCOCYTE ANTIGEN B (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CARTER V., DUNN P.P.;
RT "Identification of a novel HLA-B*07 allele.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237594; CAB40714.1; -.
FT NON_TER 1
FT NON_TER 89
SQ SEQUENCE 89 AA; 10609 MW; 659B735B CRC32;

Query Match 100.0%; Score 51; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10
DB 74 RESLNLRGY 83

RESULT 13
P79615
ID P79615 PRELIMINARY; PRT; 89 AA.
AC P79615
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MHC CLASS I HLA-B (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHOPEK M., CAO K., ZHANG G.H.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88249; AAB48493.1; -.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 89
SQ SEQUENCE 89 AA; 10489 MW; A7D3DF93 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10
DB 74 RESLNLRGY 83

RESULT 14
P79620
ID P79620 PRELIMINARY; PRT; 89 AA.
AC P79620
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MHC CLASS I HLA-B (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHOPEK M., CAO K., ZHANG G.H.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88254; AAB48498.1; -.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 89
SQ SEQUENCE 89 AA; 10501 MW; FA3EFCDC CRC32;

Query Match 100.0%; Score 51; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESLNLRGY 10
DB 74 RESLNLRGY 83

RESULT 15
O19548
ID O19548 PRELIMINARY; PRT; 89 AA.
AC O19548
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MHC CLASS I ANTIGEN (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```



OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CAO K., BURDETT L., ZHANG H., FERNANDEZ-VINA M.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016299; AAB69444.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
FT NON\_TER 89 89  
SQ SEQUENCE 89 AA; 10550 MW; 11452E40 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 89;  
Best Local Similarity 100.0%; Pred. No. 0.014;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RESLRNLRGY 10  
Db 74 RESLRNLRGY 83

Search completed: February 8, 2000, 19:16:17  
Job time: 21504 sec

**THIS PAGE BLANK (USPTO)**



```

REFERENCE 1 (bases 1 to 243)
AUTHORS Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,
Rungroung,E., Bejchandra,S., Juji,T., Tokunaga,K. and
Grosse-Wilde,H.
TITLE B15 alleles (B*1525)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 243)
AUTHORS Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,
Rungroung,E., Bejchandra,S., Juji,T., Tokunaga,K. and
Grosse-Wilde,H.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
FEATURES Location/Qualifiers
source 1..243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
exon 1..243
/gene="HLA-B"
/allele="HLA-B*1525; similar to exon 2 of B*1501"
/number=2
BASE COUNT 52 a 79 c 84 g 28 t
ORIGIN
|||||
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
195 CGAGAGAGCTCGGAACTGGCGGCTAC 224

seq_name: gb_pr2:HS1ABM1

seq_documentation_block: 243 bp DNA PRI 26-MAR-1997
LOCUS HS1ABM1
DEFINITION Human cell line THAI DCH012 MHC class I HLA-B gene (allele
HLA-B*1525), exon 2.
ACCESSION U91336
VERSION U91336.1 GI:1906679
KEYWORDS
SEGMENT 1 of 2
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 243)
AUTHORS Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,
Rungroung,E., Bejchandra,S., Juji,T., Tokunaga,K. and
Grosse-Wilde,H.
TITLE B15 alleles (B*1525)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 243)
AUTHORS Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,
Rungroung,E., Bejchandra,S., Juji,T., Tokunaga,K. and
Grosse-Wilde,H.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
FEATURES Location/Qualifiers

```

```

source 1..243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
exon 1..243
/gene="HLA-B"
/allele="HLA-B*1525; similar to exon 2 of B*1501"
/number=2
BASE COUNT 52 a 79 c 84 g 28 t
ORIGIN
|||||
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
195 CGAGAGAGCTCGGAACTGGCGGCTAC 224

seq_name: gb_pr2:AF014769

seq_documentation_block: 246 bp DNA PRI 27-AUG-1997
LOCUS AF014769
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*1502 allele),
exon 2 and partial cds.
ACCESSION AF014769
VERSION AF014769.1 GI:2345103
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 246)
AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
TITLE B15 alleles (B*1502)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 246)
AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
FEATURES Location/Qualifiers
source 1..246
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="Thai DCH4060"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
gene <1..>246
/gene="HLA-B"
/allele="HLA-B*1502"
exon 1..246
/gene="HLA-B"
/number=2
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"

```

/protein\_id="AAB67807.1"  
/db\_xref="GI:2345104"  
/translation="TAMSRGCEPRFIANGVYDDTQFVRFSDAASPRMAPRAPWIE  
QEGPEYDNTQISKNTQTYRESLNLRGYNQSEA"  
BASE COUNT 55 a 81 c 82 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x AF014769 ..

Align seg 1/1 to: AF014769 from: 1 to: 246

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
198 CGAGAGAGCTGCGGACCTGCGCGGCTAC 227

seq\_name: gb\_pr2:AF014771

seq\_documentation\_block:  
LOCUS AF014771 246 bp DNA PRI 27-AUG-1997  
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1502 allele),  
exon 2 and partial cds.

ACCESSION AF014771

VERSION AF014771.1 GI:2345107

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 246)

AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,

Rungroung,E. and Bejchandra,S.

TITLE B15 alleles (B\*1502)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 246)

AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.

TITLE Direct Submission

JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand

FEATURES

source Location/Qualifiers

1..246  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH4061"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
<1..>246  
/gene="HLA-B"  
/allele="HLA-B\*1502"  
1..246  
/gene="HLA-B"  
/note="2"  
/number=2  
/gene="HLA-B"  
/codon\_start=3  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67809.1"  
/db\_xref="GI:2345108"

gene

exon

CDS

BASE COUNT 55 a 81 c 82 g 28 t  
ORIGIN

/translation="TAMSRGCEPRFIANGVYDDTQFVRFSDAASPRMAPRAPWIE  
QEGPEYDNTQISKNTQTYRESLNLRGYNQSEA"  
BASE COUNT 55 a 81 c 82 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x AF014771 ..

Align seg 1/1 to: AF014771 from: 1 to: 246

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
198 CGAGAGAGCTGCGGACCTGCGCGGCTAC 227

seq\_name: gb\_pr2:AF014773

seq\_documentation\_block:  
LOCUS AF014773 246 bp DNA PRI 27-AUG-1997

DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1502 allele),  
exon 2 and partial cds.

ACCESSION AF014773

VERSION AF014773.1 GI:2345111

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 246)

AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,

Rungroung,E. and Bejchandra,S.

TITLE B15 alleles (B\*1502)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 246)

AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.

TITLE Direct Submission

JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand

FEATURES

source Location/Qualifiers

1..246  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH3086"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
<1..>246  
/gene="HLA-B"  
/allele="HLA-B\*1502"  
1..246  
/gene="HLA-B"  
/note="2"  
/number=2  
/gene="HLA-B"  
/codon\_start=3  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67811.1"  
/db\_xref="GI:2345112"  
/translation="TAMSRGCEPRFIANGVYDDTQFVRFSDAASPRMAPRAPWIE  
QEGPEYDNTQISKNTQTYRESLNLRGYNQSEA"  
BASE COUNT 55 a 81 c 82 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

```

US-08-653-294-34 x AF014773
Align seg 1/1 to: AF014773 from: 1 to: 246
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
|||||
198 CGAGAGAGCCTGCGGAACCTGCGCGCTAC 227

seq_name: gb_pr2:AF014775
seq_documentation_block:
LOCUS AF014775 246 bp DNA PRI 27-AUG-1997
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*1502 allele),
exon 2 and partial cds.
ACCESSION AF014775
VERSION AF014775.1 GI:2345115
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 246)
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
B15 alleles (B*1502)
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 246)
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
Direct Submission
TITLE
Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
JOURNAL
FEATURES
Location/Qualifiers
1..246
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Thai DCH018"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
<1..>246
/gene="HLA-B"
/allele="HLA-B*1502"
1..246
/gene="HLA-B"
/number=2
/note="2"
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"
/protein_id="AAB67813.1"
/db_xref="GI:2345116"
/translation="TAMSRPGRGEPRIAVGYVDDTQFVRFSDAASPRMAPRAPWIE
QEGPEYWDRTQISKNTQTYRESRLRGYNQSEA"
BASE COUNT 55 a 81 c 82 g 28 t
ORIGIN

alignment_scores:
Quality: 51.00 Length: 10
Ratio: 5.100 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-34 x AF014775
Align seg 1/1 to: AF014775 from: 1 to: 246
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
|||||
198 CGAGAGAGCCTGCGGAACCTGCGCGCTAC 227

seq_name: gb_pr2:AF014777
seq_documentation_block:
LOCUS AF014777 246 bp DNA PRI 27-AUG-1997
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*1502 allele),
exon 2 and partial cds.
ACCESSION AF014777
VERSION AF014777.1 GI:2345119
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 246)
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
B15 alleles (B*1502)
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 246)
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
Direct Submission
TITLE
Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
JOURNAL
FEATURES
Location/Qualifiers
1..246
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Thai DCH017"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
<1..>246
/gene="HLA-B"
/allele="HLA-B*1502"
1..246
/gene="HLA-B"
/number=2
/note="2"
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"
/protein_id="AAB67815.1"
/db_xref="GI:2345120"
/translation="TAMSRFGRGEPRIAVGYVDDTQFVRFSDAASPRMAPRAPWIE
QEGPEYWDRTQISKNTQTYRESRLRGYNQSEA"
BASE COUNT 55 a 81 c 82 g 28 t
ORIGIN

alignment_scores:
Quality: 51.00 Length: 10
Ratio: 5.100 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-34 x AF014777
Align seg 1/1 to: AF014777 from: 1 to: 246
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
|||||
198 CGAGAGAGCCTGCGGAACCTGCGCGCTAC 227

seq_name: gb_pr2:AF014779
seq_documentation_block:
LOCUS AF014779 246 bp DNA PRI 27-AUG-1997
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*1502 allele),
exon 2 and partial cds.

```



Rungrong,E. and Bejchandra,S.  
 Direct Submission  
 Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
 Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
 Thailand

FEATURES Location/Qualifiers  
 source 1. .246  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="Thai DCH016"  
 /chromosome="6"  
 /map="6p21"  
 /cell\_type="lymphoblastoid"

gene <1. .>246  
 /gene="HLA-B"  
 /allele="HLA-B\*1502"  
 exon 1. .246  
 /gene="HLA-B"  
 /note="2"  
 /number=2  
 CDS <1. .>246  
 /gene="HLA-B"  
 /codon\_start=3  
 /product="MHC class I antigen HLA-B"  
 /protein\_id="AAB67821.1"  
 /db\_xref="GI:2345132"

/translation="TAMSRPGCEPRFIAGVYDDTQFVRFSDAASPRMAPRPWIE  
 OEGPEYWDRTQISKNTQTYRESLNLRGYNQSEA"  
 BASE COUNT 55 a 81 c 82 g 28 t  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x AF014783 ..

Align seg 1/1 to: AF014783 from: 1 to: 246

1 ArgGlusSerLeuArgAsnLeuArgGlyTyr 10  
 |||||  
 198 CGAGAGACCTGCGGACCTGCGGGCTAC 227

seq\_name: gb\_pr2:AF014785

seq\_documentation\_block: 250 bp DNA PRI 27-AUG-1997  
 LOCUS AF014785  
 DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1525 allele),  
 exon 2 and partial cds.

ACCESSION AF014785

VERSION AF014785.1 GI:2345135

KEYWORDS human.

SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 250)

AUTHORS Chananayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungrong,E. and Bejchandra,S.

TITLE B15 alleles (B\*1525)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 250)

AUTHORS Chananayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungrong,E. and Bejchandra,S.

TITLE Direct Submission

JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
 Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
 Thailand

FEATURES Location/Qualifiers

source 1. .250

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="Thai DCH1109"  
 /chromosome="6"  
 /map="6p21"  
 /cell\_type="lymphoblastoid"  
 gene 1. .250  
 /gene="HLA-B"  
 /allele="HLA-B\*1525"  
 exon 1. .250  
 /gene="HLA-B"  
 /note="2"  
 /number=2  
 CDS <1. .>250  
 /gene="HLA-B"  
 /codon\_start=1  
 /product="MHC class I antigen HLA-B"  
 /protein\_id="AAB67823.1"  
 /db\_xref="GI:2345136"  
 /translation="FYTAMSRPGCEPRFIAGVYDDTQFVRFSDAASPRMAPRPAPW  
 IEQEGPEYWDRETQISKNTQTYRESLNLRGYNQSEA"  
 BASE COUNT 54 a 81 c 84 g 31 t  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x AF014785 ..

Align seg 1/1 to: AF014785 from: 1 to: 250

1 ArgGlusSerLeuArgAsnLeuArgGlyTyr 10  
 |||||  
 202 CGAGAGACCTGCGGACCTGCGGGCTAC 231

seq\_name: gb\_pr2:AF014787

seq\_documentation\_block: 250 bp DNA PRI 27-AUG-1997  
 LOCUS AF014787  
 DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1525 allele),  
 exon 2 and partial cds.

ACCESSION AF014787

VERSION AF014787.1 GI:2345139

KEYWORDS human.

SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 250)

AUTHORS Chananayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungrong,E. and Bejchandra,S.

TITLE B15 alleles (B\*1525)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 250)

AUTHORS Chananayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungrong,E. and Bejchandra,S.

TITLE Direct Submission

JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
 Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
 Thailand

FEATURES Location/Qualifiers

source 1. .250

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="Thai DCH3258"

/chromosome="6"

/map="6p21"

/cell\_type="lymphoblastoid"

gene 1. .250



exon  
/gene="HLA-B"  
/allele="HLA-B\*1525"  
1..250  
/gene="HLA-B"  
/note="2"  
/number=2  
<1..>250

CDS  
/gene="HLA-B"  
/codon\_start=1  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67825.1"  
/db\_xref="GI:2345140"  
/translation="FYTAMSRPGRGEPRFIAVGYVDDTQFVRFSDAASPRMAPRAPW  
IEQSGPEVDRETQISKNTQTYRESLNLRGYNQSEA"  
BASE COUNT 54 a 81 c 84 g 31 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x AF014787 ..  
Align seg 1/1 to: AF014787 from: 1 to: 250  
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
202 CGAGAGAGCGCTGCGGAACCTGCGCGCTAC 231

seq\_name: gb\_pr2:AF014789

seq\_documentation\_block:  
LOCUS AF014789 250 bp DNA PRI 27-AUG-1997  
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1525 allele),  
exon 2 and partial cds.  
ACCESSION AF014789  
VERSION AF014789.1 GI:2345143  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 250)  
AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungrong,E. and Bejchandra,S.  
TITLE B15 alleles (B\*1525)  
REFERENCE 2 (bases 1 to 250)  
AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungrong,E. and Bejchandra,S.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand  
FEATURES  
source Location/Qualifiers  
1..250  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_lines="Thai DCH025"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
1..250  
/gene="HLA-B"  
/allele="HLA-B\*1525"  
1..250  
/gene="HLA-B"  
/note="2"  
/number=2  
<1..>250

/gene="HLA-B"  
/codon\_start=1  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67827.1"  
/db\_xref="GI:2345144"  
/translation="FYTAMSRPGRGEPRFIAVGYVDDTQFVRFSDAASPRMAPRAPW  
IEQSGPEVDRETQISKNTQTYRESLNLRGYNQSEA"  
BASE COUNT 54 a 81 c 84 g 31 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x AF014789 ..  
Align seg 1/1 to: AF014789 from: 1 to: 250  
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
202 CGAGAGAGCGCTGCGGAACCTGCGCGCTAC 231

seq\_name: gb\_pr2:HSB1501N1

seq\_documentation\_block:  
LOCUS HSB1501N1 254 bp DNA PRI 10-JUL-1996  
DEFINITION Human MHC class I protein HLA-B heavy chain (B\*1501new allele)  
gene, partial exon 2.  
ACCESSION U59965  
VERSION U59965.1 GI:1399896  
KEYWORDS HLA B antigen; major histocompatibility complex.  
SEGMENT 1 of 2  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 254)  
AUTHORS Mitsuishi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUN-1996) Tissue Typing Laboratory, UCLA School of  
Medicine, 950 Veteran Ave., Los Angeles, CA 90095, USA  
FEATURES  
source Location/Qualifiers  
1..254  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="peripheral blood mononuclear cells"  
/chromosome="6"  
/map="6p21.3"  
<1..254  
/gene="HLA-B"  
/number=2

exon  
BASE COUNT 55 a 81 c 85 g 33 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x HSB1501N1 ..  
Align seg 1/1 to: HSB1501N1 from: 1 to: 254  
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
206 CGAGAGAGCGCTGCGGAACCTGCGCGCTAC 235

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-34 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Feb 8, 2000 7:31 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -DEV-rlp

-MODEL-frame+P2n.model -DEV-rlp  
-Q/cgml\_1/USPTO.spool/US08653294/runat\_04022000\_160701\_15807/app\_query.fasta.2  
-DB=N\_Geneseq\_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blom62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pcet  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=100000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

Search information block:

Query: US-08-653-294-34

Query length: 10

Database: N\_Geneseq\_36:\*

Database sequences: 311585

Database length: 125096042

Search time (sec): 873.190000

score\_list:

Sequence	Strid	Orig	Zscore	Escore	Len	Documentation
N_Geneseq_36:Q12115	+	51.00	169.52	0.0767	1089	HLA-B*35 exon. HLA-B*35 gene - u
N_Geneseq_36:Q75974	+	51.00	157.69	0.3497	4059	HLA-B*7 expression vector. New
N_Geneseq_36:Q75973	+	51.00	155.88	0.4411	4965	HLA-B*7/beta-2 microglobulin e
N_Geneseq_36:Q12116	+	44.00	143.58	2.14	1101	HLA-C exon Cb-1. HLA-C gene, D
N_Geneseq_36:Q12117	+	44.00	143.58	2.14	1101	HLA-C exon Cb-2. HLA-C gene, D
N_Geneseq_36:Q22446	+	39.00	123.46	28.18	1323	Human CPM2 DNA. New chemopro
N_Geneseq_36:Q22445	+	39.00	123.46	28.21	1324	Human CPM2 mutant DNA. New ch
N_Geneseq_36:Q12217	+	38.00	118.70	51.91	1490	Acyl-ACP-thioesterase coding s
N_Geneseq_36:Q35447	+	38.00	117.74	58.75	1659	Rat truncated rsk3 coding sequ
N_Geneseq_36:Q35448	+	38.00	117.65	59.37	1674	Truncated human hsk3 coding se
N_Geneseq_36:Q35472	+	38.00	115.10	82.38	2224	Rat rsk3 coding sequence. DNA
N_Geneseq_36:Q35473	+	38.00	114.19	92.63	2462	Human hsk3 coding sequence. DN
N_Geneseq_36:Q08414	+	38.00	113.97	95.20	2521	Human small conductance calcit
N_Geneseq_36:Q14153	+	37.00	126.48	19.13	416	Human secreted protein 5. Efr 5
N_Geneseq_36:Q30458_0	+	37.00	76.32	1.2e+04	110000	Rhizobium species plasmid p
N_Geneseq_36:Q30459_0	+	37.00	76.32	1.2e+04	110000	Rhizobium species symbiotic
N_Geneseq_36:Q11953	+	35.00	120.76	39.35	342	Anti-human Rhd BRAD-3 MAb (VH
N_Geneseq_36:Q91381	+	35.00	120.76	39.88	346	Intron 2 from human follistatin
N_Geneseq_36:Q56202	+	35.00	105.64	277.17	1858	Partial coding sequence of sal
N_Geneseq_36:Q90126	+	35.00	105.64	277.17	1858	Plasmid pCK1 SAPP cDNA insert.
N_Geneseq_36:Q38418	+	35.00	105.64	277.17	1858	pCK1 cDNA insert encoding sal
N_Geneseq_36:Q94624	+	35.00	99.84	583.35	3542	Brush-1 cDNA probe. New cDNA s
N_Geneseq_36:Q15158	+	35.00	98.12	736.84	4286	CDK4 cyclin-dependent kinase-
N_Geneseq_36:Q34625	+	35.00	97.97	741.13	4359	Brush-1 cDNA. New cDNA sequen
N_Geneseq_36:Q33449	+	35.00	97.44	793.74	4626	Oryza sativa L. pistil specif
N_Geneseq_36:Q35528	+	34.00	113.84	96.79	495	Human secreted protein gene 18
N_Geneseq_36:Q1608	+	34.00	107.13	228.91	1044	Insert of pCCA10 encoding carb
N_Geneseq_36:Q52673	+	34.00	105.57	279.67	1242	HTIe receptor DNA. Stably co-t
N_Geneseq_36:Q47076	+	34.00	104.68	313.43	1371	B. canis 21B4/rhoptv antigen
N_Geneseq_36:Q01731	+	34.00	100.81	514.80	2108	Mouse liver cancer-originate
N_Geneseq_36:Q43711	+	34.00	99.41	616.03	2463	Sequence of the human gene 5-H
N_Geneseq_36:Q13139	+	34.00	85.96	3.5e+03	10996	Enterococcus faecalis genome
N_Geneseq_36:Q13401	+	34.00	85.93	3.5e+03	11036	Enterococcus faecalis genome
N_Geneseq_36:Q21210	+	34.00	70.94	2.3e+04	59407	Methanococcus jannaschii larg
N_Geneseq_36:Q20248_07	+	34.00	65.25	4.7e+04	110000	Continuation (8 of 10) of
N_Geneseq_36:Q194815	+	33.00	139.96	3.40	18	Human leukocyte antigen class I
N_Geneseq_36:Q42475	+	33.00	123.93	26.56	107	Huntl coding sequence. Huntl an
N_Geneseq_36:Q72980	+	33.00	120.03	43.77	165	Huntingtin gene antisense produ
N_Geneseq_36:Q23425	+	33.00	107.02	232.16	701	Mouse TWRL3 DNA. New Tumor Necr
N_Geneseq_36:Q69757	+	33.00	106.47	249.05	745	Isopentenyl transferase ipt gen
N_Geneseq_36:Q18559	+	33.00	102.43	418.34	1168	Mus musculus tumour necrosis d

N\_Geneseq\_36:V43032 + 33.00 101.43 475.43 1305 Streptococcus pneumoniae po  
N\_Geneseq\_36:V32620 + 33.00 101.11 495.66 1353 Genomic DNA sequence of Str  
N\_Geneseq\_36:Q80003 - 33.00 98.73 672.62 1763 ADP-glucose-pyrophosphoryla  
N\_Geneseq\_36:V29686 + 33.00 95.40 1.0e+03 2553 A. thermophilum thermostabi  
seq\_name: N\_Geneseq\_36:Q12115  
seq\_documentation\_block:  
ID Q12115 standard; DNA; 1089 BP.  
AC Q12115;  
DT 29-AUG-1991 (first entry)  
DE HLA-B\*35 exon.  
KW Human leukocyte antigen; probe; major histocompatibility complex;  
KW MHC; class I; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 1..1089  
FT FT /\*tag= a  
PN J03112486-A.  
PD 14-MAY-1991.  
PF 22-SEP-1989; 247697.  
PR 22-SEP-1989; JP-247697.  
PA (OLYU ) OLYMPUS OPTICAL KK.  
DR WPI; 91-182991/25.  
DR P-PSDB; R12464.  
PT HLA-B\*35 gene - used in DNA probe and transformant cells for  
PT immunising animals, for developing monoclonal antibody.  
PS Claim 1; Page 1; lipp; Japanese.  
CC Probes comprising part of the sequence can be used to identify  
CC Class I genes. The DNA can be expressed for immunisation of  
CC animals and prodn. of monoclonal antibodies specific for the  
CC HLA-B\*35 antigen. See also J03112485 and J03112487.  
SQ Sequence 1089 BP; 221 A; 336 C; 359 G; 173 T;

alignment\_scores:

Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x Q12115 ..  
Align seg 1/1 to: Q12115 from: 1 to: 1089  
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
295 CGAGAGAGCTCGGAACCTCGCGCTAC 324

seq\_name: N\_Geneseq\_36:Q75974

seq\_documentation\_block:

ID Q75974 standard; cDNA; 4059 BP.  
AC Q75974;  
DT 23-AUG-1995 (first entry)  
DE PHLA-B\*7 expression vector.  
KW Expression vector; PHLA-B\*7; heavy human HLA-B\*7; bicistronic mRNA;  
KW light beta-2 microglobulin; class I major histocompatibility complex;  
KW MHC; human leukocyte antigen; HLA; covalently closed circular DNA; ds.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_feature 1..354  
FT FT /\*tag= a  
FT FT /\*note= "PB322 backbone contg. bacterial origin of  
FT FT replication"  
FT FT 355..1170  
FT FT /\*tag= b  
FT FT /\*note= "kanamycin resistance gene open reading frame;  
FT FT the gene is taken from the transposable element  
FT FT Tn903"  
FT FT poly\_a\_signal  
FT FT complement (1410..1177)  
FT FT /\*tag= c  
FT FT /\*note= "SV40 polyA signal sequence"

```

FT intron complement (1412..1560)
FT /tag= d
FT /note= "SV40 small t intron"
FT 3'utr complement (1561..1794)
FT /tag= e
FT /note= "3' untranslated region of HLA-B7
FT heavy chain mRNA"
FT complement (1795..2880)
FT /tag= f
FT /note= "HLA-B7 open reading frame"
FT ltr complement (2886..3415)
FT /tag= g
FT /note= "Rous sarcoma virus 3' LTR promoter region"
FT 3416..4059
FT /tag= h
FT /note= "pBR322 backbone"
FT WO9429469-A.
FT 22-DEC-1994.
FT 27-MAY-1994; U06069.
FT 07-JUN-1993; US-074344.
FT (UNMI ) UNIV MICHIGAN.
FT PA Lew D, Marquet M, Nabel EG, Nabel GJ;
FT WPI; 95-036494/05.
FT New vectors for gene therapy, partic for tumours - comprising
FT genetic material encoding one or more cistron(s) which express
FT immunogenic or therapeutic peptide(s)
FT PS Claim 9; Page 42-43; 50pp; English.
FT CC This HLA-B7 antigen encoding plasmid was developed to incorporate many
CC advantageous features, eg. the kanamycin resistance gene. The
CC eradication of two open reading frames encoding portions of SV40 viral
CC as a cassette into which cistrons may be inserted and removed at will
CC proteins lowers the risk of tumorigenicity. The vector may also operate
CC for the transcription and subsequent translation of peptides of interest.
CC The vector is used partic. for the treatment of neoplastic disease,
CC eg. melanoma, and provides enhanced gene delivery and expression
CC in vivo.
FT Sequence 4059 BP; 975 A; 1051 C; 1033 G; 1000 T;

alignment_scores:
  Quality: 51.00 Length: 10
  Ratio: 5.100 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-34 x Q75974/rev ..
Align seg 1/1 to reverse of: Q75974 from: 1 to: 4059

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
|||||
2589 CGAGAGAGCGCTGCGGACCTGCGGCGGTAC 2560

seq_name: N_Geneseq_36:Q75973
seq_documentation_block:
ID Q75973 standard; cDNA; 4965 BP.
AC Q75973;
DT 23-AUG-1995 (first entry)
DE pHLA-B7/beta-2 microglobulin expression vector.
KW expression vector; pHLA-B7/beta-2 microglobulin; heavy human HLA-B7;
KW light beta-2 microglobulin; class I major histocompatibility complex;
KW MHC; bicistronic mRNA; human leukocyte antigen; HLA;
KW covalently closed circular DNA; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT ltr 1..529
FT /tag= a
FT /note= "Rous sarcoma virus LTR promoter domain, derived
FT for the Schmidt-Rupin strain nucleotides
FT 8673-9146. This region also includes a 56 bp
FT region of a synthetic oligonucleotide which

```

```

FT modifies this regulatory sequence to effect a
FT higher level of expression of downstream
FT sequences. The oligonucleotide removes a
FT polyadenylation signal sequence originally found
FT in the RSV DNA sequence."
FT 531..534
FT /tag= b
FT /label= consensus_Kozak_signal_sequence
FT 535..1620
FT /tag= c
FT /note= "HLA-B7 heavy chain open reading frame"
FT 535..606
FT /tag= d
FT /note= "encodes putative signal peptide of the
FT HLA-B7 heavy chain"
FT 607..1620
FT /tag= e
FT /note= "encodes putative HLA-B7 heavy chain mature
FT peptide"
FT 1621..1853
FT /tag= f
FT /note= "3' untranslated sequence of HLA-B7 heavy
FT chain mRNA"
FT 1854..1888
FT /tag= g
FT /note= "multiple cloning site, forms a junction
FT between the HLA-B7 sequence and the EMCV-CITE
FT sequence, and is used to facilitate subcloning"
FT 1889..2479
FT /tag= h
FT /note= "murine encephalomyocarditis CAP-independent
FT translational enhancer (EMCV-CITE); taken from
FT nucleotides 253-843 of cloned EMCV genomic DNA.
FT It is a non-coding regulatory sequence functioning
FT as an internal entry point for the eukaryotic
FT ribosomal subunits when located within a mRNA
FT mol. It enables the translational start codon of
FT the beta-2 microglobulin, downstream of the HLA-B7
FT stop codon on this bicistronic mRNA to be
FT recognised by the ribosome"
FT 2480..2839
FT /tag= i
FT /note= "encodes beta-2 microglobulin; this cDNA is
FT deriv. from chimpanzee (differs to the human
FT cDNA by only 4 bases)"
FT 2840..2846
FT /tag= j
FT /note= "3' untranslated region of the beta-2
FT microglobulin mRNA"
FT 2847..2870
FT /tag= k
FT /note= "synthetic linker"
FT 3'UTR ..3111
FT /tag= l
FT /note= "bovine growth hormone 3'UTR and transcriptional
FT terminator; it starts at a blunt-ended BglI site
FT within the 3'UTR of the mRNA coding sequence"
FT 2979..2984
FT /tag= m
FT 3112..3151
FT /tag= n
FT /note= "synthetic linker to facilitate cloning"
FT complement (3151..3967)
FT /tag= o
FT /note= "kanamycin resistance gene open reading frame;
FT the gene is taken from the transposable element
FT Tn903"
FT 4014..4965
FT /tag= p
FT /note= "pBR322 backbone contg. bacterial origin of
FT replication, it represents nucleotides 2244-3193"
FT WO9429469-A.
FT 22-DEC-1994.

```



```

FT      WO9859055-A2.
PN      30-DEC-1998.
PD      24-JUN-1998; D01796.
PF      24-JUN-1997; DE-026823.
PR      (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA      Pram C, Schwab M;
PI      WPI: 99-081282/07.
DR      P-PSDB: W93087.
PT      New chemoprotector regulated modifier proteins - useful in the
FT      detoxification of metabolites, particularly aflatoxins
PS      Claim 2; Fig 8; 31pp; German.
CC      This sequence encodes a novel protein designated CPRM2 (chemoprotector
CC      regulated modifier) that inhibits the effects of genotoxic and/or
CC      cytotoxic substances. The invention describes the isolation of CPRM1,
CC      CPRM2, CPRM3, mCPRM1, allele s, and mCPRM1, allele r. Transformants
CC      containing such proteins are used to produce recombinant proteins and
CC      nucleic acids for use in gene therapy and to raise antibodies which are
CC      used to (i) detoxify geno- and/or cyto-toxic metabolites and (ii) detect
CC      presence of CPRM proteins (or nucleic acid encoding them, including
CC      detection of mutations indicative of increased risk of cancer, e.g.
CC      prenatally). The proteins can be used to eliminate carcinogenic
CC      aflatoxins and to treat neurodegeneration, e.g. Alzheimer's disease,
CC      associated with production of malondialdehyde, a toxic product or the
CC      cyclo-oxygenase reaction and prostaglandin metabolism.
SQ      Sequence 1323 BP; 331 A; 368 C; 338 G; 286 T;

alignment_scores:
Quality: 39.00 Length: 10
Ratio: 4.333 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-34 x X22446/rev ..
Align seg 1/1 to reverse of: X22446 from: 1 to: 1323

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
:::|||||:::||||| |||||||
1058 AAAGAAGTGTGAGAACAAAGAAGGTTAC 1029

seq_name: N_Geneseq_36:X22450

seq_documentation_block:
ID X22450 standard; DNA; 1324 BP.
AC X22450;
DT 20-MAY-1999 (first entry)
DE Human CPRM2 mutant DNA.
KW CPRM1; mouse; chemoprotector regulated modifier; genotoxin; cytotoxin;
KW CPRM2; CPRM3; mCPRM1; allele s; allele r; gene therapy; antibody;
KW detoxification; detection; mutant; cancer; carcinogenic; aflatoxin;
KW treatment; neurodegeneration; Alzheimer's disease; malondialdehyde;
KW cyclo-oxygenase reaction; prostaglandin metabolism; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 74..223
FT /*tag= a
FT /product= "CPRM2 mutant"

PN      WO9859055-A2.
PD      30-DEC-1998.
PF      24-JUN-1998; D01796.
PR      24-JUN-1997; DE-026823.
PA      (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI      Pram C, Schwab M;
DR      WPI: 99-081282/07.
PT      New chemoprotector regulated modifier proteins - useful in the
FT      detoxification of metabolites, particularly aflatoxins
PS      Disclosure; Fig 12; 31pp; German.
CC      This sequence encodes a novel protein designated CPRM2 which is a mutant

```

```

CC      chemoprotector regulated modifier that inhibits the effects of
CC      genotoxic and/or cytotoxic substances. The invention describes the
CC      isolation of CPRM1, CPRM2, CPRM3, mCPRM1, allele s, and mCPRM1,
CC      allele r. Transformants containing such proteins are used to produce
CC      recombinant proteins and nucleic acids for use in gene therapy and to
CC      raise antibodies which are used to (i) detoxify geno- and/or cyto-toxic
CC      metabolites and (ii) detect presence of CPRM proteins (or nucleic acid
CC      encoding them, including detection of mutations indicative of increased
CC      risk of cancer, e.g. prenatally). The proteins can be used to eliminate
CC      carcinogenic aflatoxins and to treat neurodegeneration, e.g. Alzheimer's
CC      disease, associated with production of malondialdehyde, a toxic product
CC      or the cyclo-oxygenase reaction and prostaglandin metabolism.
SQ      Sequence 1324 BP; 332 A; 368 C; 338 G; 286 T;

alignment_scores:
Quality: 39.00 Length: 10
Ratio: 4.333 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-34 x X22450/rev ..
Align seg 1/1 to reverse of: X22450 from: 1 to: 1324

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
:::|||||:::||||| |||||||
1059 AAAGAAGTGTGAGAACAAAGAAGGTTAC 1030

seq_name: N_Geneseq_36:T32217

seq_documentation_block:
ID T32217 standard; cDNA to mRNA; 1490 BP.
AC T32217;
DT 07-FEB-1997 (first entry)
DE Acyl-ACP-thioesterase coding sequence.
KW Acyl-ACP-thioesterase; AAT; cuphea leptopoda; aliphatic acid synthesis;
KW middle chain aliphatic acid; ss.
OS Cuphea leptopoda hemsl.
FH Key Location/Qualifiers
FT cds 30..1289
FT /*tag= a
FT /product= Acyl-ACP-thioesterase

PN      J08173165-A.
PD      09-JUL-1996.
PF      22-DEC-1994; 320547.
PR      22-DEC-1994; JP-320547.
PA      (NORQ ) NORINSUISANSO CHUGOKU NOGYO SHIKENBACH.
DR      WPI: 96-365584/37.
DR      P-PSDB: W02020.
PT      DNA encoding Cuphea leptopoda Hemsl. acyl-ACP-thio:esterase - for
PT      synthesis of middle chain aliphatic acids, useful in pharmaceutical
PT      preps, and foods.
PS      Claim 2; Page 6; 7pp; Japanese.
CC      This sequence represents the coding sequence for the Cuphea leptopoda
CC      acyl-ACP-thioesterase (AAT). This sequence was isolated by synthesizing
CC      a AAT cDNA, and using that sequence to probe for a fragment of the AAT
CC      gene (see T32217). The isolated fragment was then used to isolate the
CC      AAT gene, which was then amplified. AAT is a key enzyme in the synthesis
CC      of middle chain aliphatic acids, and as such is useful in pharmaceutical
CC      preparations, and foods.
SQ      Sequence 1490 BP; 351 A; 332 C; 402 G; 405 T;

alignment_scores:
Quality: 38.00 Length: 9
Ratio: 4.222 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 66.667

alignment_block:
US-08-653-294-34 x T32217/rev ..
Align seg 1/1 to reverse of: T32217 from: 1 to: 1490

```

```

2 GluSerLeuArgAsnLeuArgGlyTyr 10
  ::::::::::::::::::::::::::::::
1472 AAATCAATCGGACCAICGAGGATAC 1446

seq_name: N_Geneseq_36:V35447

seq_documentation_block:
ID V35447 standard; CDNA; 1659 BP.
AC V35447;
DT 01-OCT-1998 (first entry)
DE Rat truncated RSK3 coding sequence.
KW Small conductance calcium-activated potassium channel protein 1;
KW RSK3; rat; potassium ion flux; ss.
OS Rattus sp.
PN WO9811139-A1.
PD 19-MAR-1998.
PF 10-SEP-1997; U16033.
PR 17-APR-1997; US-045233.
PR 11-SEP-1996; US-026451.
PR 07-MAR-1997; US-040052.
PA (ICAG-) ICAGEN INC.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Adelman JP, Bond CT, Maylie J, Silvia CP;
DR WPI; 98-207332/18.
DR P-PSDB; W63703.
PT DNA encoding calcium-activated potassium channel - useful in assays
PT to identify compounds which increase or decrease potassium ion flux
PS Claim 3; Page 105; 151pp; English.
CC This sequence encodes the small conductance calcium-activated
CC potassium channel protein 3 (SK3) of the invention. The proteins of the
CC invention are monomers of a calcium-activated potassium channel, where
CC the monomer: (i) has a calculated molecular weight of between 40 and
CC 80 kDa; and (ii) has a unit conductance of between 2 and 60 pS when the
CC monomer is in the functional polymeric form of a potassium channel and is
CC expressed in a Xenopus oocyte. Antibodies specific for the protein, and
CC probes specific for the DNA can be used to detect the presence of the
CC protein or DNA sequences in a sample. Host cells expression of the
CC protein can be used in assays to identify compounds which increase or
CC decrease the potassium ion flux through the protein. The transfected host
CC cell can also be used for the recombinant production of the protein. The
CC DNA sequences can also be used for determine mutations in the SK and IK
CC genes in a computer system. The proteins encoded by the SK and IK genes
CC can be used in a computer system for determining their three dimensional
CC structure, which is useful for determining ligands that bind to the
CC proteins.
SQ Sequence 1659 BP; 404 A; 499 C; 398 G; 358 T;

alignment_scores:
  Quality: 38.00 Length: 10
  Ratio: 3.800 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-34 x V35447/rev ..
Align seg 1/1 to reverse of: V35447 from: 1 to: 1659

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
  ::::::::::::::::::::::::::::::
62 CGAGAGCGCTGAGCGCGTGGGGGCTTC 33

seq_name: N_Geneseq_36:V35458

seq_documentation_block:
ID V35458 standard; CDNA; 1674 BP.
AC V35458;
DT 01-OCT-1998 (first entry)
DE Truncated human hSK3 coding sequence.
KW Small conductance calcium-activated potassium channel protein 3;
KW hSK3; human; potassium ion flux; ss.
OS Homo sapiens.

alignment_scores:
  Quality: 38.00 Length: 10
  Ratio: 3.800 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-34 x V35447/rev ..
Align seg 1/1 to reverse of: V35447 from: 1 to: 1659

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
  ::::::::::::::::::::::::::::::
62 CGAGAGCGCTGAGCGCGTGGGGGCTTC 33

seq_name: N_Geneseq_36:V35458

seq_documentation_block:
ID V35458 standard; CDNA; 1674 BP.
AC V35458;
DT 01-OCT-1998 (first entry)
DE Truncated human hSK3 coding sequence.
KW Small conductance calcium-activated potassium channel protein 3;
KW hSK3; human; potassium ion flux; ss.
OS Homo sapiens.

alignment_scores:
  Quality: 38.00 Length: 10
  Ratio: 3.800 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-34 x V35458/rev ..
Align seg 1/1 to reverse of: V35458 from: 1 to: 1674

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
  ::::::::::::::::::::::::::::::
62 CGGAGGCGCTGAGCGCGTGGGGGCTTC 33

seq_name: N_Geneseq_36:V35472

seq_documentation_block:
ID V35472 standard; CDNA; 2224 BP.
AC V35472;
DT 01-OCT-1998 (first entry)
DE Rat RSK3 coding sequence.
KW Small conductance calcium-activated potassium channel protein 3;
KW RSK3; rat; potassium ion flux; ss.
OS Rattus sp.
FH Key Location/Qualifiers
FT CDS 2..2200
FT /tag= a
PN WO9811139-A1.
PD 19-MAR-1998.
PF 10-SEP-1997; U16033.
PR 17-APR-1997; US-045233.
PR 11-SEP-1996; US-026451.
PR 07-MAR-1997; US-040052.
PA (ICAG-) ICAGEN INC.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Adelman JP, Bond CT, Maylie J, Silvia CP;
DR WPI; 98-207332/18.
DR P-PSDB; W63715.
PT DNA encoding calcium-activated potassium channel - useful in assays

```







```
FT FT      /*tag= k
FT FT      /standard_name= "ORF K11"
FT FT      /product= "protein of unknown function"
FT FT      /note= "homologous to the FixU gene"
FT FT      complement (434107. .434433)
FT FT      /*tag= l
FT FT      /standard_name= "ORF K12"
FT FT      /product= "protein of unknown function"
FT FT      complement (434517. .434711)
FT FT      /*tag= m
FT FT      /standard_name= "ORF K13"
FT FT      /product= "ferrioxin/ferrioxin-like protein"
FT FT      /note= "homologous to the FdxN gene"
FT FT      complement (434753. .436234)
FT FT      /*tag= n
FT FT      /standard_name= "ORF K14"
FT FT      /gene= "nifB"
FT FT      /product= "protein involved in Femo co-factor
FT FT      biosynthesis"
FT FT      complement (436460. .438130)
FT FT      /*tag= o
FT FT      /standard_name= "ORF K15"
FT FT      /gene= "nifA"
FT FT      /product= "positive regulator of nif, fix and other
FT FT      genes"
FT FT      complement (438297. .438590)
FT FT      /*tag= p
FT FT      /standard_name= "ORF K16"
FT FT      /gene= "fixX"
FT FT      /product= "protein required for nitrogenase activity"
FT FT      complement (438605. .439912)
FT FT      /*tag= q
FT FT      /standard_name= "ORF K17"
FT FT      /gene= "fixC"
FT FT      /product= "protein required for nitrogenase activity"
FT FT      complement (439923. .441032)
FT FT      /*tag= r
FT FT      /standard_name= "ORF K18"
FT FT      /gene= "fixB"
FT FT      /product= "protein required for nitrogenase activity"
FT FT      complement (441042. .441899)
FT FT      /*tag= s
FT FT      /standard_name= "ORF K19"
FT FT      /gene= "fixA"
FT FT      /product= "protein required for nitrogenase activity"
FT FT      complement (442316. .442636)
FT FT      /*tag= t
FT FT      /standard_name= "ORF K20"
FT FT      /product= "protein of unknown function"
FT FT      complement (443313. .443879)
FT FT      /*tag= u
FT FT      /standard_name= "ORF K21"
FT FT      /product= "protein of unknown function"
FT FT      444337. .445029
FT FT      /*tag= v
FT FT      /standard_name= "ORF K22"
FT FT      /product= "ferrioxin-like protein"
FT FT      /note= "homologous to the NifQ gene"
FT FT      445088. .446602
FT FT      /*tag= w
FT FT      /standard_name= "ORF K23"
FT FT      /gene= "dctA"
FT FT      /product= "Cd-dicarboxylate transport protein"
FT FT      /note= "homologous to the DctAI gene"
FT FT      446599. .447843
FT FT      /*tag= x
FT FT      /standard_name= "ORF L1"
FT FT      /product= "cytochrome P450-like protein"
FT FT      /note= "homologous to the CamC gene"
FT FT      447844. .448500
FT FT      /*tag= y
FT FT      /standard_name= "ORF L2"
FT FT      /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT FT      protein"
FT FT      448497. .450203
FT FT      /*tag= z
FT FT      /standard_name= "ORF L3"
FT FT      /product= "putative protein with degradative function"
FT FT      450341. .451396
FT FT      /*tag= aa
FT FT      /standard_name= "ORF L4"
FT FT      /product= "luciferase alpha-subunit-like protein"
FT FT      /note= "homologous to the LuxA gene"
FT FT      452980. .454494
FT FT      /*tag= ab
FT FT      /standard_name= "ORF L6"
FT FT      /gene= "nifD"
FT FT      /product= "alpha-subunit of Femo protein of nitrogenase"
FT FT      454590. .456131
FT FT      /*tag= ac
FT FT      /standard_name= "ORF L7"
FT FT      /gene= "nifK"
FT FT      /product= "beta-subunit of Femo protein of nitrogenase"
FT FT      456187. .457677
FT FT      /*tag= ad
FT FT      /standard_name= "ORF L8"
FT FT      /product= "protein involved in Femo co-factor
FT FT      biosynthesis"
FT FT      /note= "homologous to the NifB gene"
FT FT      457687. .459096
FT FT      /*tag= ae
FT FT      /standard_name= "ORF L9"
FT FT      /product= "protein involved in Femo co-factor
FT FT      biosynthesis"
FT FT      /note= "homologous to the FixF gene"
FT FT      459093. .459575
FT FT      /*tag= af
FT FT      /standard_name= "ORF L10"
FT FT      /product= "protein of unknown function"
FT FT      /note= "homologous to the NifX gene"
FT FT      459579. .460067
FT FT      /*tag= ag
FT FT      /standard_name= "ORF L11"
FT FT      /product= "protein of unknown function"
FT FT      460501. .460920
FT FT      /*tag= ah
FT FT      /standard_name= "ORF L12"
FT FT      /product= "protein similar to part of the Fe protein
FT FT      of nitrogenase"
FT FT      /note= "homologous to the NifH gene"
FT FT      461228. .461545
FT FT      /*tag= ai
FT FT      /standard_name= "ORF L13"
FT FT      /product= "protein of unknown function"
FT FT      463201. .464739
FT FT      /*tag= aj
FT FT      /standard_name= "ORF L14"
FT FT      /product= "peptidase-like protein"
FT FT      /note= "homologous to the bi-MPP gene"
FT FT      464736. .466073
FT FT      /*tag= ak
FT FT      /standard_name= "ORF L15"
FT FT      /product= "processing protease-like protein"
FT FT      466073. .466073

alignment_scores:
Quality: 37.00 Length: 9
Ratio: 4.625 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 88.889

alignment_block:
US-08-653-294-34 x V30458_0 ..
Align seg 1/1 to: V30458_0 from: 1 to: 110000

1 ArgGluSerLeuArgAsnLeuArgGly 9
```

Wed Feb 9 08:50:31 2000

us-08-653-294-34.rng

Page 9

|||||  
15658 CGCGAAAGTCTTAGGAATCTCTTGGG 15684

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-34 to: EST:\* out\_format : pfs

Date: Feb 8, 2000 6:23 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet p2n.model -DEV=xlp  
-o=/cnl1/USPTO.spool/US08653294/runat\_04022000.160700.15770/app\_query.fasta.2  
-DB=EST -QFWT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -GAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFWT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-34

Query length: 10

Database: EST\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 7600.090000

score\_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
gb_est12:AA319533	+	51.00	194.48	103	AA319533 EST21772 Adrenal gland
gb_est13:AA361477	+	51.00	188.68	189	AA361477 EST70761 T-cell lymph
gb_est13:AA325603	+	51.00	187.31	218	AA325603 EST60621 Activated T-c
gb_est12:AA294911	+	51.00	185.66	259	AA294911 EST100074 Pancreas tum
gb_est13:AA322960	+	51.00	185.26	270	AA322960 EST1101 Activated T-c
gb_est8:AA058454	+	51.00	182.17	373	AA058454 z167d07.r1 Stratagene
gb_est11:AA224068	+	51.00	181.58	397	AA224068 z11f02.r1 Stratagene
gb_est10:AA160317	+	51.00	180.39	405	AA160317 z056c07.r1 Stratagene
gb_est7:AA0489	+	51.00	180.88	427	AA0489 zc84b01.r1 Pancreatic Is
gb_est12:AA310808	+	51.00	180.53	443	AA310808 EST181593 Jurkat T-cell
gb_est9:CI8310	+	51.00	179.86	475	CI8310 CI8310 Human placenta ch
gb_est17:AA663896	+	51.00	179.78	479	AA663896 ae74d01.sl Stratagene
gb_est35:AL039796	+	51.00	179.31	503	AL039796 DKF2p43B1912.r1 434
gb_est11:AA263135	+	51.00	176.02	710	AA263135 PMV0598 KGI-a Lambda Z
gb_est27:AA124623	-	44.00	161.23	209	AA124623 tf34g01.x1 NCI CGAP_B
gb_est23:AA1097437	-	44.00	160.03	237	AA1097437 qb61e08.x1 NCI CGAP_B
gb_est14:AA449607	+	44.00	156.03	360	AA449607 z06h02.r1 Soares tota
gb_est11:Y55088	+	44.00	156.00	361	Y55088 yb45c09.r1 Stratagene fe
gb_est11:AA227277	+	44.00	155.64	375	AA227277 zr20h03.r1 Stratagene
gb_est31:AI1706797	+	44.00	154.81	409	AI1706797 UI-R-AA1-aab-h-09-0-UI
gb_est31:AI1706862	+	44.00	154.67	415	AI1706862 UI-R-AA1-aac-c-04-0-UI
gb_est23:AA109801	-	44.00	154.22	435	AA109801 qall1a09.x1 NCI CGAP_B
gb_est11:AA243206	-	44.00	154.13	439	AA243206 qy2d12.x1 NCI CGAP_B
gb_est15:AA477506	+	44.00	153.50	469	AA477506 z035f05.sl Soares ovar
gb_est4:HA23327	+	44.00	153.30	479	HA23327 ym52e08.r1 Soares infant
gb_est26:AA849373	+	44.00	153.18	485	AA849373 EST192140 Normalized r
gb_est26:AA924062	+	44.00	152.72	509	AA924062 UI-R-A1-da-e-07-0-UI.s
gb_est37:AI939506	+	44.00	152.72	509	AI939506 qy2d12.x5 NCI CGAP_B
gb_est26:AA945159	+	44.00	152.30	532	AA945159 EST229271 Normalized r
gb_est26:AA945159	+	44.00	152.01	548	AA945159 EST200658 Normalized r
gb_est26:AI232284	+	44.00	150.41	648	AI232284 EST228972 Normalized r
gb_est10:AA248102	-	42.00	144.36	287	AA248102 HS-2015.AL.H09 MR CIT
gb_est27:AA140530	+	41.00	146.52	297	AA140530 EST238823 Normalized r
gb_est29:AA1602357	+	41.00	145.14	343	AA1602357 UI-R-AA0-w-h-05-0-UI.s
gb_est9:AA100680	+	41.00	143.23	326	AA100680 z19b02.r1 Stratagene
gb_est38:AA060987	-	41.00	141.73	390	AA060987 UI-R-BH1-ann-c-01-0-UI
gb_est9:AA083156	+	40.00	141.70	396	AA083156 zn08d04.r1 Stratagene
gb_est9:AA121088	+	40.00	139.63	411	AA121088 zm2d05.r1 Stratagene
gb_gss11:AAQ33000	-	40.00	138.21	477	AAQ33000 HS-5003.A2.A08.r7 RPCH
gb_est10:AA158109	-	40.00	137.85	62.13	AA158109 zo57h10.sl Stratagene
gb_est10:AA160714	+	40.00	137.22	70.66	AA160714 zo77e11.r1 Stratagene

gb\_gss13:AAQ468916 - 40.00 137.02 72.47 540 ! AQ468916 HS-5140.B2.H09.r7A  
gb\_est11:AA203261 - 40.00 132.13 135.60 900 ! AA203261 zx55d11.r1 Soares f  
gb\_est4:FA14616 + 39.00 143.90 29.97 177 ! F14616 SSO4H02 Porcine small

seq\_name: gb\_est12:AA319533

seq\_documentation\_block:

LOGUS AA319533 103 bp mRNA EST 19-APR-1997  
DEFINITION EST21772 Adrenal gland tumor Homo sapiens cDNA 5' end similar to  
similar to major histocompatibility complex, class I, B  
(GB:M16102), mRNA sequence.

ACCESSION AA319533  
VERSION AA319533.1 GI:1971860  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 103)

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.B., Geoghagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.

TITLE

Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

JOURNAL

Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE

12140200

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1393672.  
Other ESTs: THC169519  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699036  
Fax: 3018699423

FEATURES

Location/Qualifiers  
1..103  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):120083"  
/db\_xref="taxon:9606"  
/clone\_lib="Adrenal gland tumor"  
/dev\_stage="adult"  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M3 Reverse

BASE COUNT

26 a 35 c 29 g 13 t  
ORIGIN  
alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-653-294-34 x AA319533

Align seg 1/1 to: AA319533 from: 1 to: 103

1 ArgGlusSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
27 CGAGAGAGCCTCGGAGACCTTCGCGGCTAC 56

seq\_name: gb\_est13:AA361477

seq\_documentation\_block: 189 bp mRNA EST 21-APR-1997  
LOCUS AA361477 T-cell lymphoma Homo sapiens cDNA 5' end similar to  
DEFINITION EST70761 similar to major histocompatibility complex, class I, Bw62.3, mRNA  
sequence.  
VERSION AA361477.1 GI:2013795  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 189)  
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

TITLE Nature 377 (6547 Suppl), 3-174 (1995)  
JOURNAL 12140200  
MEDLINE  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404737.  
Other ESTs: THCI69519

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
Location/Qualifiers  
Source  
1..189  
/organism="Homo sapiens"  
/db\_xref="Arcc (inhost):165623"  
/db\_xref="taxon:9606"  
/clone\_lib="T-cell lymphoma"  
/cell\_type="T-lymphocyte"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 45 a 60 c 55 g 25 t 4 others  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x AA361477

Align seg 1/1 to: AA361477 from: 1 to: 189

1 ArgGlusSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
37 CGAGAGAGCCTCGGAGACCTTCGCGGCTAC 66

seq\_name: gb\_est13:AA352603

seq\_documentation\_block: 218 bp mRNA EST 21-APR-1997  
LOCUS AA352603 Activated T-cells XX Homo sapiens cDNA 5' end similar to  
DEFINITION EST60621 similar to major histocompatibility complex, class I, Bw62.3, mRNA  
sequence.

ACCESSION AA352603  
VERSION AA352603.1 GI:2004923  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 218)

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

TITLE Nature 377 (6547 Suppl), 3-174 (1995)  
JOURNAL 12140200  
MEDLINE  
COMMENT On May 8, 1995 this sequence version replaced gi:800964.  
Other ESTs: THCI72938

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
Location/Qualifiers  
Source  
1..218  
/organism="Homo sapiens"  
/db\_xref="Arcc (inhost):152802"  
/db\_xref="taxon:9606"  
/clone\_lib="Activated T-cells XX"  
/cell\_type="T-lymphocyte"  
/dev\_stage="adult"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 52 a 70 c 62 g 28 t 6 others  
ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-34 x AA352603 ..  
 Align seg 1/1 to: AA352603 from: 1 to: 218

1 ArgGlusertLeuArgAsnLeuArgGlyTyr 10  
 |||||  
 39 CGAGAGAGCTCGGGAACCTGCGCGGTAC 68

seq\_name: gb\_est12:AA294911

seq\_documentation\_block: 259 bp mRNA EST 18-APR-1997  
 LOCUS AA294911  
 DEFINITION EST100074 Pancreas tumor I Homo sapiens cDNA 5' end similar to  
 similar to major histocompatibility complex, class I, Bw62.3, mRNA  
 sequence.

ACCESSION AA294911

VERSION AA294911.1 GI:1947266

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 259)

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
 Glodek,A., Gnehm,C.L., Hanna,M.C., Fritchman,J.L., Geoghagen,N.S.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
 Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
 Kozak,D.L., Kunsch,C., Hungjun,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Raymond,L.P., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
 Fraser,C.M. and Venter,J.C.

Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

## JOURNAL

MEDLINE

COMMENT

On Nov 29, 1993 this sequence version replaced gi:430148.

Other ESTs: THCL72938

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..259

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):190413"

/db\_xref="taxon:9606"

/clone\_lib="Pancreas tumor I"

/dev\_stage="adult"

/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI"

60 a 87 c 75 g 30 t 7 others

BASE COUNT

## ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-34 x AA294911 ..

Align seg 1/1 to: AA294911 from: 1 to: 259

1 ArgGlusertLeuArgAsnLeuArgGlyTyr 10  
 |||||  
 42 CGAGAGAGCTCGGGAACCTGCGCGGTAC 71

seq\_name: gb\_est13:AA352960

seq\_documentation\_block: 270 bp mRNA EST 21-APR-1997

LOCUS AA352960

DEFINITION EST11101 Activated T-cells XX Homo sapiens cDNA 5' end similar to

similar to major histocompatibility complex, class I, B61, mRNA

sequence.

ACCESSION AA352960

VERSION AA352960.1 GI:2005353

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 270)

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
 Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
 Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
 Kozak,D.L., Kunsch,C., Hungjun,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Raymond,L.P., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
 Fraser,C.M. and Venter,J.C.

Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

12140200

On May 8, 1995 this sequence version replaced gi:801269.

Other ESTs: THCL72938

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..270

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):153240"

/db\_xref="taxon:9606"

/clone\_lib="Activated T-cells XX"

/cell\_type="T-lymphocyte"

FEATURES

source

/dev\_stage="adult"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 62 a 80 c 88 g 35 t 5 others  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x AA352960 ..

Align seg 1/1 to: AA352960 from: 1 to: 270

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
182 CGAGAGAGCTCGGAACCTCGCGGCTAC 211

seq\_name: gb\_est8:AA058454

seq\_documentation\_block:

LOCUS AA058454 373 bp mRNA EST 01-DEC-1996  
DEFINITION z167d07.r1 Stratagene colon (#937204) Homo sapiens cDNA clone  
IMAGE:509677 5' similar to gb:M24039.cdsl HLA CLASS I  
HISTOCOMPATIBILITY ANTIGEN, B-18 B\*1801 ALPHA (HUMAN);, mRNA  
sequence.

ACCESSION AA058454  
VERSION AA058454.1 GI:1551280  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
and Marra, M.

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

On Sep 12, 1996 this sequence version replaced gi:1393357.

Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1592 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 268.

FEATURES

source  
1..373  
/organism="Homo sapiens"  
/db\_xref="GDB:3813009"  
/db\_xref="taxon:9606"  
/clone="IMAGE:509677"

/clone\_lib="Stratagene colon (#937204)"

/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: colon; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. T-84 colonic epithelial cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTT 3'.

BASE COUNT 73 a 119 c 116 g 57 t 8 others  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x AA058454 ..

Align seg 1/1 to: AA058454 from: 1 to: 373

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
229 CGAGAGAGCTCGGAACCTCGCGGCTAC 258

seq\_name: gb\_est11:AA224068

seq\_documentation\_block:

LOCUS AA224068 397 bp mRNA EST 19-FEB-1997  
DEFINITION z111f02.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone  
IMAGE:648507 5' similar to gb:L22649.cdsl HLA CLASS I  
HISTOCOMPATIBILITY ANTIGEN, B-39 B\*3901 ALPHA (HUMAN);, mRNA  
sequence.

ACCESSION AA224068  
VERSION AA224068.1 GI:1844610  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
and Marra, M.

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

On Nov 29, 1993 this sequence version replaced gi:430429.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev1 ET from Amersham

High quality sequence stop: 295.

FEATURES

source  
1..397  
/organism="Homo sapiens"  
/db\_xref="GDB:5588245"  
/db\_xref="taxon:9606"  
/clone="IMAGE:648507"  
/clone\_lib="Stratagene hNT neuron (#937233)"

/dev\_stage="hNT neurons"

/lab\_host="SOLR (kanamycin resistant)"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dt.

Differentiated, post mitotic hNT neurons. Average insert  
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
GAATTCGCACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTT 3'.

BASE COUNT 77 a 129 c 131 g 56 t 4 others

ORIGIN



alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-653-294-34 x AA224068 ..

Align seg 1/1 to: AA224068 from: 1 to: 397

1 ArgGlusLeuArgAsnLeuArgGlyTyr 10  
 ||||||||||||||||||||||||||||||||  
 297 CGAGAGAGCGTCGNAACCTCGCGGCTAC 326

seq\_name: gb\_est10:AA160317

seq\_documentation\_block:  
 LOCUS AA160317 405 bp mRNA EST 09-MAR-1998  
 DEFINITION z056c07.r1 Stragatene pancreas (#937208) Homo sapiens cDNA clone  
 IMAGE:590892 5' similar to gb:U11265 HLA CLASS I HISTOCOMPATIBILITY  
 ANTIGEN, B-35 B\*3501 ALPHA (HUMAN);, mRNA sequence.

ACCESSION AA160317  
 VERSION AA160317.1 GI:1734956  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 405)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1406818.

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1671 Std Error: 0.00  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 332.

FEATURES  
 source  
 1..405  
 /organism="Homo sapiens"  
 /db\_xref="GDB:4622278"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:590892"  
 /clone\_lib="Stragatene pancreas (#937208)"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. Pancreatic adenocarcinoma cell line. Average  
 insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor  
 sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'  
 CTCGAGTGTGTTTTTTTTTT 3'"

BASE COUNT 80 a 132 c 132 g 58 t 3 others  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-653-294-34 x AA160317 ..

Align seg 1/1 to: AA160317 from: 1 to: 405

1 ArgGlusLeuArgAsnLeuArgGlyTyr 10  
 ||||||||||||||||||||||||||||||||  
 281 CGAGAGAGCGTCGGAACCTCGCGGCTAC 310

seq\_name: gb\_est7:W40489

seq\_documentation\_block:  
 LOCUS W40489 427 bp mRNA EST 20-MAY-1996  
 DEFINITION zc84b01.r1 Pancreatic Islet Homo sapiens cDNA clone IMAGE:328969 5'  
 similar to gb:U11265 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35  
 B\*3501 ALPHA (HUMAN);, mRNA sequence.

ACCESSION W40489  
 VERSION W40489.1 GI:1324496  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 427)  
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thiery, Meg J.,  
 Trevasaki, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.  
 and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478

On Jan 25, 1995 this sequence version replaced gi:637865.

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 397.

FEATURES  
 source  
 1..427  
 Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="GDB:1263173"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:328969"  
 /clone\_lib="Pancreatic Islet"  
 /tissue\_type="pancreatic islet"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
 EcoRI; Site 2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)  
 Takeda et al. Cloned unidirectionally. Primer: Oligo dt.  
 -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
 sequence: 5' CTCGAGTGTGTTTTTTTTTT 3'"

BASE COUNT 87 a 141 c 134 g 62 t 3 others  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-653-294-34 x W40489 ..

Align seg 1/1 to: W40489 from: 1 to: 427

1 ArgGlusLeuArgAsnLeuArgGlyTyr 10  
 ||||||||||||||||||||||||||||||||

279 CGAGAGCCTCGGAACCTCGCGGGCTAC 308

seq\_name: gb\_est12:AA310808

seq\_documentation\_block: 443 bp mRNA EST 19-APR-1997  
 LOCUS AA310808  
 DEFINITION EST181593 Jurkat T-cells v Homo sapiens cDNA 5' end similar to similar to major histocompatibility complex, class I, B0704, mRNA sequence.  
 ACCESSION AA310808  
 VERSION AA310808.1 GI:1963136  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 443)  
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
 TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
 MEDLINE 12140200  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397854.  
 Other ESTs: THC180721  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlavage@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 Location/Qualifiers  
 1..443  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):156811"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Jurkat T-cells v"  
 /cell\_type="T-lymphocyte"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI."

BASE COUNT 86 a 144 c 143 g 64 t 6 others  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x AA310808

Align seg 1/1 to: AA310808 from: 1 to: 443

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
 |||||  
 292 CGAGAGCCTCGGAACCTCGCGGGCTAC 321

seq\_name: gb\_est9:C18310

seq\_documentation\_block: 475 bp mRNA EST 02-OCT-1996  
 LOCUS C18310  
 DEFINITION C18310 Human placenta cDNA (Tfujiiwara) Homo sapiens cDNA clone GEN-560D07 5', mRNA sequence.

ACCESSION C18310  
 VERSION C18310.1 GI:1579912  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 475)  
 AUTHORS Fujiwara,T., Hirano,H., Hishigaki,H., Horie,M., Kawai,A., Kuga,Y., Kyushiki,H., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Suzuki,M., Takaichi,A., Takeda,S., Watanabe,T., Maekawa,H., Nakamura,Y. and Takahashi,E.

TITLE Otsuka cDNA project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393837.  
 Contact: Tsutomu Fujiwara  
 Otsuka GEN Research Institute  
 Otsuka Pharmaceutical Co.,Ltd  
 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan  
 Tel: 0886-65-2888  
 Fax: 0886-37-1035.

FEATURES  
 Location/Qualifiers  
 1..475  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="GEN-560D07"  
 /clone\_lib="Human placenta cDNA (Tfujiiwara)"  
 /tissue\_type="placenta"

BASE COUNT 89 a 161 c 154 g 68 t 3 others  
 ORIGIN

alignment\_scores:

Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x C18310

Align seg 1/1 to: C18310 from: 1 to: 475

1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10  
 |||||  
 299 CGAGAGCCTCGGAACCTCGCGGGCTAC 328

seq\_name: gb\_est17:AA663896

seq\_documentation\_block:

LOCUS AA663896 479 bp mRNA EST 15-DEC-1997  
 DEFINITION ae74d01.s1 Stragene schizo brain S11 Homo sapiens cDNA clone IMAGE:969889 3', similar to gb:M28203 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-62 B\*1504 ALPHA (HUMAN);, mRNA sequence.

ACCESSION AA663896  
 VERSION AA663896.1 GI:2617887  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 479)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R. WashU-NCI human EST Project Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1394858.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: polyt not found  
Seq primer: -40m13 fwd. ET from Amerham  
High quality sequence stop: 430.

#### FEATURES

source  
1. 479  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:969889"  
/clone\_lib="Stratagene schizo brain Sil"  
/sex="male"  
/tissue\_type="schizophrenic brain S-11 frontal lobe"  
/dev\_stage="34 years old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Aleutian. Schizophrenic suicide. Random primed into EcoRI site of ZAP II Vector. Mass excised. Custom library. Avg insert length 1.4kb. Material obtained by Johnston N., Torrey, E.F., Volken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of individuals with psychiatric Diseases (Unpublished) Stanley Neurovirology Laboratory, Johns Hopkins School of Medicine, Baltimore MD."  
BASE COUNT 91 a 165 c 158 g 64 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-34 x AA663896 ..

Align seg 1/1 to: AA663896 from: 1 to: 479

1 ArgGlusSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
283 CGAGAGAGCTGCGGACCTGCGGCGGTAC 312

seq\_name: gb\_est35:AL039796

seq\_documentation\_block:  
LOCUS AL039796 503 bp mRNA EST 29-SEP-1999  
DEFINITION DKF2p434B1912\_r1 434 (synonym: htes3) Homo sapiens cDNA clone  
DKF2p434B1912 5', mRNA sequence.

ACCESSION AL039796  
VERSION AL039796.1 GI:5408804  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 503)  
AUTHORS Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

TITLE EST (Duesterhoeft, et al.)  
JOURNAL Unpublished (1999)

#### COMMENT

On Jun 2, 1999 this sequence version replaced gi:4967270.  
Contact: Duesterhoeft A

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.

This clone is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

#### FEATURES

source  
1. 503  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKF2p434B1912"  
/clone\_lib="434 (synonym: htes3)"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"  
BASE COUNT 96 a 172 c 165 g 68 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-34 x AL039796 ..

Align seg 1/1 to: AL039796 from: 1 to: 503

1 ArgGlusSerLeuArgAsnLeuArgGlyTyr 10  
|||||  
300 CGAGAGAGCTGCGGACCTGCGGCGGTAC 329

seq\_name: gb\_est11:AA263135

seq\_documentation\_block:  
LOCUS AA263135 710 bp mRNA EST 02-JUL-1998  
DEFINITION PMY0598 KGI-a Lambda Zap Express cDNA library Homo sapiens cDNA 5', mRNA sequence.

ACCESSION AA263135  
VERSION AA263135.1 GI:1898941  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 710)  
AUTHORS Claudio, J.O., Liew, C.C., Dempsey, A.A., Cukerman, E., Stewart, A.K., Na, E., Atkins, H.I., Iscove, N.N. and Hawley, R.G.

TITLE Identification of sequence-tagged transcripts differentially expressed within the human hematopoietic hierarchy

JOURNAL Genomics 50 (1), 44-52 (1998)  
MEDLINE 98292493  
COMMENT On May 5, 1995 this sequence version replaced gi:797738.

Contact: Hawley RG  
Oncology Research Laboratories  
The Toronto Hospital  
CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada  
Tel: 416 3403834  
Fax: 416 3403453  
Email: r.hawley@utoronto.ca  
Clone was randomly picked from KGIa primary library.  
Seq primer: 5' GAATTAACCTCACTAAGGG 3'  
High quality sequence stop: 710.  
Location/Qualifiers

#### FEATURES

```

source
1. 710
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Kcl-a Lambda Zap Express cDNA library"
/cell_type="promyeloblast"
/cell_line="Kcl-a"
/notes="Vector: Lambda Zap Express (Stratagene); Site_1:
EcoRI; Site_2: XhoI; Unidirectional cloning sites:
EcoRI-XhoI-mRNA was purified from Kcl-a cell line, cDNA
adaptors were ligated, followed by digestion with XhoI for
directional cloning into predigested Lambda Zap Express"

BASE COUNT      150 a   224 c   227 g   109 t
ORIGIN

alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-34 x AA263135 ..
Align seg 1/1 to: AA263135 from: 1 to: 710
1 ArgGluSerLeuArgAsnLeuArgGlyTyr 10
|||||
300 CGAGAGAGCTCGCGAACCTCGCGGCTAC 329

seq_name: gb_est27:AI424623

seq_documentation_block:
LOCUS      AI424623      209 bp      mRNA      EST      30-MAR-1999
DEFINITION  tf34901.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2098128 3',
            mRNA sequence.
ACCESSION  AI424623
VERSION    AI424623.1 GI:4270554
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 209)
            NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGA), Tumor Gene Index
JOURNAL   Unpublished (1998)
COMMENT   On Apr 21, 1998 this sequence version replaced gi:3072294.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone Distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1183 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 208.
Location/Qualifiers
1. .209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2098128"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"

```

```

/lab_host="DH10B"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCCGATATCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT      63 a   55 c   33 g   57 t   1 others
ORIGIN

alignment_scores:
  Quality: 44.00      Length: 9
  Ratio: 4.889      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-34 x AI424623/rev ..
Align seg 1/1 to reverse of: AI424623 from: 1 to: 209
1 ArgGluSerLeuArgAsnLeuArgGly 9
|||||
157 AGAGAGTCCCTTAGGAACTTGAGAGGG 131

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 04:05:44 ; Search time 133.56 Seconds  
(without alignments)  
3.547 Million cell updates/sec

Title: US-08-653-294-35

Perfect score: 102

Sequence: 1 YGRNLSRSLRLRGY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102	100.0	20	R92913	HLA-B7 CTL modul
2	102	100.0	20	R95415	HLA-B7.84-75-84 Pa
3	102	100.0	20	W33790	Peptide B7.84-75-77
4	102	100.0	20	W33797	Peptide B7.84-75-77
5	51	50.0	10	R41209	Peptide fragment o
6	51	50.0	10	R83061	HLA-B7 CTL modul
7	51	50.0	10	W07515	T-cell modulating
8	51	50.0	10	W33786	Peptide B7.75-84 t
9	51	50.0	10	W33796	Peptide B7.75-84 t
10	51	50.0	25	R41207	Peptide fragment o
11	51	50.0	25	R83073	HLA-Bw62 CTL modul
12	51	50.0	25	R95431	HLA-B7.60-84. Comp
13	51	50.0	25	R95419	HLA-Bw62.60-84. Co
14	51	50.0	362	R12454	HLA-B35 antigen. H
15	47.5	46.6	605	W62838	Glycine max antiml
16	46	45.1	509	W26792	Mouse TIE-2 recept
17	44	43.1	20	R92910	HLA-B2702 CTL modu
18	44	43.1	20	W33793	Peptide B2702.84-7
19	44	43.1	25	R41206	Peptide fragment o
20	44	43.1	25	R83072	HLA-Bw46 CTL modul
21	44	43.1	25	R95418	HLA-Bw46.60-84. Co
22	44	43.1	366	R12465	HLA-C exon Cb-1. H
23	44	43.1	366	R12466	HLA-C exon Cb-2. H
24	44	43.1	366	Y07033	Breast cancer asso
25	43	42.2	430	R20642	Glyphosate oxidore
26	43	42.2	431	R22262	Glyphosate oxidore
27	42	41.2	20	R92908	HLA-B2702 CTL modu
28	42	41.2	20	R92909	Peptide B2702 CTL modu
29	42	41.2	20	W33791	Peptide B2702.84-7
30	42	41.2	20	W33792	Peptide B2702.84-7
31	42	41.2	2233	W48711	HPV-3 JS isolate
32	42	41.2	2233	W48712	HPV-3 FRH1 cp45 v
33	42	41.2	2233	W48713	HPV-3 vero cp45 v
34	41	40.2	217	W45424	Human macrophage s

## RESULT 1

R92913  
ID R92913 standard; peptide; 20 AA.  
AC R92913;  
DT 16-MAY-1996 (first entry)  
DE HLA-B7 CTL modulating peptide (B7.84-75/75-84).  
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
KW Immunosuppressant; graft versus host disorder; transplantation; therapy;  
KW class I MHC; HLA-B7.  
OS Synthetic.  
PN W09526979-A1.  
PD 12-OCT-1995.  
PF 05-APR-1995; U04349.  
PR 05-APR-1994; US-222851.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM, Parham P;  
DR WPI: 95-358582/45.  
PT Extension of acceptance period of transplants from MHC unmatched  
PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
PT host  
PS Example 15; Page 36; 80pp; English.  
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
CC class I major histocompatibility complex (MHC) antigens. This sequence  
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
CC I MHC HLA-B7. These sequences can be used to extend the period of  
CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
CC The peptides are administered to a patient in conjunction with a  
CC subtherapeutic amount of an immunosuppressant. This is administered to  
CC the patient for a limited period of time (compared to the lifetime  
CC administration for current treatments). The peptides particularly  
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
CC of the patient.  
SQ Sequence 20 AA;

Query Match 100.0% Score 102; DB 1; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.3e-09; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0;

Qy 1 YGRNLSRSLRLRGY 20

Db 1 YGRNLSRSLRLRGY 20

## RESULT 2

R95415  
ID R95415 standard; peptide; 20 AA.  
AC R95415;  
DT 12-NOV-1996 (first entry)  
DE HLA-B7.84-75-84 Palindrome.  
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN W09513288-A1.  
PD 18-MAY-1995.

35 41 40.2 219 1 W98326 H. pylori GHPO 144  
36 41 40.2 274 1 P80911 Consensus sequence  
37 41 40.2 705 1 R66597 Human L5/3 tumour  
38 41 40.2 705 1 R66598 Human L5/3 tumour  
39 41 40.2 705 1 W14266 Human L5/3 partial  
40 41 40.2 705 1 W14267 Human L5/3 partial  
41 41 40.2 710 1 W07691 Macrophage stimula  
42 41 40.2 711 1 R66602 Human L5/3 tumour  
43 41 40.2 711 1 W07692 Macrophage stimula  
44 41 40.2 711 1 W14270 Human growth facto  
45 41 40.2 711 1 W82789 Human MSP protein.

## ALIGNMENTS

PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Compsns. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example: Page 18; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B\*7.84-75/75-84 palindromic. These sequences can be used to isolate  
 CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane  
 CC protein associated with T-cell activation in mammalian T-cells, and is  
 CC also immunologically cross reactive with the heat shock protein Hsc70.  
 CC p74 is found in a limited number of cell types, but is particularly  
 CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
 CC cell with an amphoteric detergent, and then passed through an affinity  
 CC column containing a covalently bound HLA-B\*2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B\*2702-60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytotoxicity. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 20 AA;

Query Match 100.0%; Score 102; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YGRNLSRERESLNRLGY 20  
 DB 1 YGRNLSRERESLNRLGY 20

## RESULT 3

ID W33790 standard; peptide; 20 AA.  
 AC W33790;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.84-75/75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09744351-AL.  
 PD 27-NOV-1997.  
 PF 22-MAY-1996; US-653294.  
 PR (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PA Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes, be  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.

SQ Sequence 20 AA;  
 Query Match 100.0%; Score 102; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YGRNLSRERESLNRLGY 20  
 DB 1 YGRNLSRERESLNRLGY 20

CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.

SQ Sequence 20 AA;

Query Match 100.0%; Score 102; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YGRNLSRERESLNRLGY 20  
 DB 1 YGRNLSRERESLNRLGY 20

## RESULT 4

ID W33797 standard; peptide; 20 AA.  
 AC W33797;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.84-75/75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09744351-AL.  
 PD 27-NOV-1997.  
 PF 22-MAY-1996; US-653294.  
 PR (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PA Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes, be  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.

SQ Sequence 20 AA;

Query Match 100.0%; Score 102; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YGRNLSRERESLNRLGY 20  
 DB 1 YGRNLSRERESLNRLGY 20

RESULT 5

R41209 ID R41209 standard; peptide; 10 AA.

AC R41209.

DT 15-MAR-1994 (first entry)

DE Peptide fragment of Class I HLA peptide.

KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;

KW parasitic disease; cytotoxic T lymphocyte; modulation.

OS Synthetic.

PN W09317699-A.

PD 16-SEP-1993.

PF 25-FEB-1993; U01758.

PR 02-MAR-1992; US-844716.

PA (STRD ) UNIV LELAND STANFORD JUNIOR.

PI Clayberger CA, Krensky AM;

DR WPI; 93-303134/38.

PT New peptide(s) based on Class I HLA antigen domains - used for

PT modulating cytotoxic T-lymphocyte activity towards targets

PS Claim 11; Page 54; 61pp; English.

CC The peptide is used to modulate cytotoxic T-lymphocyte (CTL)

CC activity, either by inhibition or stimulation. It can be used

CC for inhibiting CTL toxicity in transplantations, for inducing CTL

CC activity in parasitic diseases and neoplasia and in studies on viral

CC infection. The peptide can also be used for identifying CTLs which

CC bind to it and removing subsets of CTLs from a T-cell composition.

CC This peptide sequence is more commonly found within larger peptide

CC compounds of not more than 30 amino acids in length.

CC Sequence 10 AA;

SQ

Query Match 50.0%; Score 51; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.043;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20

DB 1 RESLNRLRGY 10

RESULT 6

R83061 ID R83061 standard; peptide; 10 AA.

AC R83061;

DT 15-MAY-1996 (first entry)

DE HLA-B7 CTL modulating peptide (B7.75-84).

KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;

KW immunosuppressant; graft versus host disorder; transplantation; therapy;

KW Class I MHC; HLA-B7.

OS Synthetic.

PN W09526979-A1.

PD 12-OCT-1995.

PF 05-APR-1995; U04349.

PR 05-APR-1994; US-222851.

PA (STRD ) UNIV LELAND STANFORD JUNIOR.

PI Clayberger C, Krensky AM, Parham P;

DR WPI; 95-358582/46.

PT Extension of acceptance period of transplants from MHC unmatched

PT donor hosts - using Class I B75-84 MHC antigen of the recipient

PT host

PS Claim 13; Page 66; 80pp; English.

CC This sequence represents a fragment of a class I major histocompatibility

CC complex (MHC) antigen. This sequence corresponds to residues 75-84 of

CC the alpha-1 domain of the class I MHC HLA-B7. This sequence, and the

CC peptide fragments represented by R83062-R83085, R83090-R83096 and

CC R83097-R83093 can be used to extend the period of acceptance by a

CC recipient of a transplant from an MHC unmatched donor. The peptides are

CC administered to a patient in conjunction with a subtherapeutic amount of

CC immunosuppressant. This is administered to the patient for a limited

CC period of time (compared to the lifetime administration for current

CC treatments). The peptides particularly modulate (or inhibit) the

CC activity of the cytotoxic T lymphocytes (CTLs) of the patient.

CC Sequence 10 AA;

SQ

Query Match 50.0%; Score 51; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.043;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20

DB 1 RESLNRLRGY 10

Query Match 50.0%; Score 51; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.043;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20

DB 1 RESLNRLRGY 10

RESULT 7

W07515 ID W07515 standard; peptide; 10 AA.

AC W07515;

DT 04-AUG-1997 (first entry)

DE T-cell modulating peptide #4.

KW T-cell modulator; autoimmune disease; tissue destruction; alpha-domain;

KW mammal; major histocompatibility complex; MHC class I; antigen; perforin;

KW insulin-dependent diabetes mellitus; multiple sclerosis; inflammation;

KW rheumatoid arthritis; psoriasis; pemphigus vulgaris; Sjogren's disease;

KW thyroid disease; Hashimoto's thyroiditis; myasthenia gravis; granzyme;

KW autologous target cell; cytokine release; T cell activation; therapy.

OS Synthetic.

PN W09635443-A1.

PD 14-NOV-1996.

PF 05-APR-1996; U04710.

PR 12-MAY-1995; US-440504.

PA (SANG-) SANGSTAT MEDICAL CORP.

PI Buelow R;

DR WPI; 96-518410/51.

PT Treatment of auto-immune disease by admin. of peptide(s) corresp. to

PT major histocompatibility complex antigens - esp. for delaying onset

PT of clinical symptoms of insulin dependent diabetes by modulating T

PT cell mediated attack on target cells

PS Claim 7; Page 20; 24pp; English.

CC W07512-W07518 represent T-cell modulating peptides that can be used in

CC the method of the invention. These sequences are based on a portion of

CC the generic peptide corresponding to residues 70-91 of the alpha-1 domain

CC of the major histocompatibility complex (MHC) class I antigen (see

CC W07510). The method is for affecting the course of an autoimmune disease

CC involving T-cell mediated destruction of tissue in mammals. These

CC peptides are used especially to treat insulin-dependent diabetes

CC mellitus, preferably being administered during the pre-clinical stage to

CC delay onset of the disease. Other diseases that can be treated are

CC multiple sclerosis, rheumatoid arthritis, psoriasis, pemphigus vulgaris,

CC Sjogren's disease, thyroid disease, Hashimoto's thyroiditis, myasthenia

CC gravis, etc. The peptides modulate T-cell mediated attack on autologous

CC target cells, and may also reduce inflammation, swelling, and release of

CC cytokines, perforins, granzymes etc. associated with T cell activation.

CC Sequence 10 AA;

SQ

Query Match 50.0%; Score 51; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.043;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20

DB 1 RESLNRLRGY 10

RESULT 8

W33786 ID W33786 standard; peptide; 10 AA.

AC W33786;

DT 19-JUN-1998 (first entry)

DE Peptide B7.75-84 tested for immunomodulating activity.

KW immunomodulating dimer; immunosuppressant drug; CTL activation;

KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;

KW rejection.

OS Synthetic.

PN W09744351-A1.

PD 27-NOV-1997; U08689.  
 PF 22-MAY-1997; US-653294.  
 PR (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI; 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W3784-98 and W3778-9 were assayed for their immunomodulating  
 activity. A peptide-type compound or variant is claimed which has  
 immunomodulating activity, including the N-terminal acylated and/or  
 C-terminal amidated or esterified forms of up to 60 amino acids, where  
 the peptide-type compound comprises the formula: A-B, where A, B =  
 (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 acid. The sequence in the brackets may optionally be absent or truncated  
 at any peptide type bond within the brackets. The compounds comprise  
 amino acid sequences related to a Class I HLA-B alpha domain (positions  
 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 undesirably attacking cells in a host or in vitro. They can also be  
 used in combination with antigenic peptides or proteins of interest to  
 activate CTLs. They can also inhibit the proliferation of T cells in  
 response to anti-CD3. The peptide can be used for preventing rejection  
 of transplants or for treating autoimmune diseases, e.g. diabetes,  
 rheumatoid arthritis and lupus erythematosus. The products can also be  
 used for detection and diagnosis.  
 SQ Sequence 10 AA;

Query Match 50.0%; Score 51; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 |||||  
 DB 1 RESLNRLRGY 10

RESULT 9  
 W37796  
 ID W37796 standard; peptide; 10 AA.  
 AC W37796;  
 DT 19-JUN-1998 (first entry)  
 DE Peptide B7.75-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09744351-Al.  
 PD 27-NOV-1997.  
 PR 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI; 98-086530/08.

PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W3784-98 and W3778-9 were assayed for their immunomodulating  
 activity. A peptide-type compound or variant is claimed which has  
 immunomodulating activity, including the N-terminal acylated and/or  
 C-terminal amidated or esterified forms of up to 60 amino acids, where  
 the peptide-type compound comprises the formula: A-B, where A, B =  
 (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 acid. The sequence in the brackets may optionally be absent or truncated  
 at any peptide type bond within the brackets. The compounds comprise

CC amino acid sequences related to a Class I HLA-B alpha domain (positions  
 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 undesirably attacking cells in a host or in vitro. They can also be  
 used in combination with antigenic peptides or proteins of interest to  
 activate CTLs. They can also inhibit the proliferation of T cells in  
 response to anti-CD3. The peptide can be used for preventing rejection  
 of transplants or for treating autoimmune diseases, e.g. diabetes,  
 rheumatoid arthritis and lupus erythematosus. The products can also be  
 used for detection and diagnosis.  
 SQ Sequence 10 AA;

Query Match 50.0%; Score 51; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 |||||  
 DB 1 RESLNRLRGY 10

RESULT 10  
 R41207  
 ID R41207 standard; peptide; 25 AA.  
 AC R41207;  
 DT 15-MAR-1994 (first entry)  
 DE Peptide fragment of Class I HLA peptide.  
 KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
 KW parasitic disease; cytotoxic T lymphocyte; modulation.  
 OS Synthetic.  
 PN W09317699-A.  
 PD 16-SEP-1993.  
 PF 25-FEB-1993; U01758.  
 PR 02-MAR-1992; US-844716.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger CA, Krensky AM;  
 DR WPI; 93-303134/38.

PT New peptide(s) based on Class I HLA antigen domains - used for  
 modulating cytotoxic T-lymphocyte activity towards targets  
 Claim 10; Page 54; 61pp; English.  
 CC The peptide (or a fragment of at least 10 amino acids, joined at at  
 least one terminus to a sequence other than that of wild type HLA  
 antigen) is used to modulate cytotoxic T-lymphocyte (CTL) activity,  
 either by inhibition or stimulation. It can be used for  
 inhibiting CTL toxicity in transplantations, for inducing CTL  
 activity in parasitic diseases and neoplasia and in studies on viral  
 infection. The peptide can also be used for identifying CTLs which  
 bind to it and removing subsets of CTLs from a T-cell composition.  
 SQ Sequence 25 AA;

Query Match 50.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 |||||  
 DB 16 RESLNRLRGY 25

RESULT 11  
 R83073  
 ID R83073 standard; peptide; 25 AA.  
 AC R83073;  
 DT 16-MAY-1996 (first entry)  
 DE HLA-Bw62 CTL modulating peptide (Bw62.60-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW Class I MHC; HLA-Bw62.  
 OS Synthetic.  
 PN W09526979-Al.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.



PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI: 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B7-84 MHC antigen of the recipient  
 PT host  
 PS Example 13; Page 32; 80pp; English.  
 CC R3061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC corresponds to residues 60-84 of the alpha-1 domain of the class I MHC  
 CC HLA-B\*62. These sequences can be used to extend the period of acceptance  
 CC by a recipient of a transplant from an MHC unmatched donor. The peptides  
 CC are administered to a patient in conjunction with a subtherapeutic amount  
 CC of an immunosuppressant. This is administered to the patient for a  
 CC limited period of time (compared to the lifetime administration for  
 CC current treatments). The peptides particularly modulate (or inhibit) the  
 CC activity of the cytotoxic T lymphocytes (CTLs) of the patient.  
 SQ Sequence 25 AA;

Query Match 50.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 DB 16 RESLNRLRGY 25

## RESULT 12

ID R95431 standard; peptide; 25 AA.  
 AC R95431;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B\*60-84.  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Compens. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 12; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B\*60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 50.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 DB 16 RESLNRLRGY 25

## RESULT 13

ID R95419 standard; peptide; 25 AA.  
 AC R95419;  
 DT 12-NOV-1996 (first entry)  
 DE HLA-B\*62.60-84.  
 KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PF 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM;  
 DR WPI: 95-194027/25.  
 PT Compens. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 9; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-B\*62.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 50.0%; Score 51; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 DB 16 RESLNRLRGY 25

## RESULT 14

ID R12464 standard; Protein; 362 AA.  
 AC R12464;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-B\*35 antigen.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I.  
 OS Homo sapiens.  
 PN J03112486-A.

PD 14-MAY-1991.  
 PF 22-SEP-1989; 247697.  
 PR 22-SEP-1989; JP-247697.  
 PA (OLIU ) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182991/25.  
 DR N-PSDB; Q12115.  
 PT HLA-B35 gene - used in DNA probe and transformant cells for  
 PR immunising animals, for developing monoclonal antibody.  
 PS Claim 1; Page 1; lipp; Japanese.  
 CC Probes comprising part of the sequence encoding this sequence can  
 be used to identify Class I genes. The DNA can be expressed for  
 CC immunisation of animals and prodn. of monoclonal antibodies  
 CC specific for the HLA-B35 antigen. See also J03112485 and J03112487.  
 SQ Sequence 362 AA;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRGY 20  
 DB 99 RESLRNLRGY 108

## RESULT 15

ID W62838 standard; Protein; 605 AA.  
 AC W52838;  
 DT 27-OCT-1998 (first entry)  
 DE Glycine max antimicrobial protein.  
 KW antimicrobial protein; infestation; control.  
 OS Glycine max  
 PN W09827805-A1.  
 PD 02-JUL-1998.  
 PF 22-DEC-1997; AU0874.  
 PR 20-DEC-1996; AU-004275.  
 PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
 PI Bower NJ, Goulter KC, Green JL, Manners JM, Marcus JP;  
 DR WPI: 98-377279/32  
 PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
 PT useful for controlling microbial infestations of plants or mammals  
 PS Claim 1; Page 63-65; 96pp; English.  
 CC The sequence is that of an antimicrobial protein which can  
 CC be used to control microbial infestations in plants and mammalian  
 CC animals.  
 SQ Sequence 605 AA;

Query Match 46.6%; Score 47.5; DB 1; Length 605;  
 Best Local Similarity 47.8%; Pred. No. 11;  
 Matches 11; Conservative 3; Mismatches 6; Indels 3; Gaps 1;

QY 1 YGR---LNRLSERRESLRNLRGY 20  
 DB 211 YGRIRVLORENQSPQLNLRDY 233

Search completed: February 8, 2000, 04:05:44  
 Job time: 9361 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 18:04:41 ; Search time 111.22 seconds  
(without alignments)  
8.482 Million cell updates/sec

Title: US-08-653-294-35

Perfect score: 102

Sequence: 1 YGRLNRLSERRESRLRGY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR\_62:\*

1: pirl:\*

2: pirl2:\*

3: pirl3:\*

4: pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	51	50.0	137	2 I38875	MHC class I antige
2	51	50.0	137	2 I38876	MHC class I antige
3	51	50.0	137	2 I38860	MHC class I antige
4	51	50.0	137	2 I38874	MHC class I antige
5	51	50.0	181	2 I59188	MHC cell surface g
6	51	50.0	270	1 HLH040	MHC class I histoc
7	51	50.0	274	2 I68774	MHC HLA-B39 chain
8	51	50.0	274	2 S24439	class I histocompa
9	51	50.0	300	2 I68701	cell surface antig
10	51	50.0	308	2 I36956	MHC class I antige
11	51	50.0	330	2 I68747	MHC class I antige
12	51	50.0	330	2 I54308	MHC HLA B71 - huma
13	51	50.0	334	2 S24436	class I histocompa
14	51	50.0	354	2 S24437	class I histocompa
15	51	50.0	354	2 S24438	class I histocompa
16	51	50.0	354	2 S24440	class I histocompa
17	51	50.0	354	2 S24433	class I histocompa
18	51	50.0	358	2 S03538	class I histocompa
19	51	50.0	361	2 I54418	MHC class I histoc
20	51	50.0	362	1 HLH0B7	MHC class I histoc
21	51	50.0	362	2 S77966	MHC class I histoc
22	51	50.0	362	2 G01230	MHC class I histoc
23	51	50.0	362	2 S16789	class I histocompa
24	51	50.0	362	2 A45880	MHC class I histoc
25	51	50.0	362	2 I37519	MHC class I histoc
26	51	50.0	362	2 I37520	MHC class I histoc
27	51	50.0	362	2 I37522	MHC class I histoc
28	51	50.0	362	2 I84431	MHC class I histoc
29	51	50.0	362	2 I72755	HLA-B*5602 - human
30	51	50.0	362	2 I84488	lymphocyte antigen

31 51 50.0 362 2 I72752 HLA-B\*5501 - human  
32 51 50.0 362 2 I72753 HLA-B\*5502 - human  
33 51 50.0 362 2 I72754 HLA-B\*5601 - human  
34 51 50.0 362 2 I38437 MHC class I histoc  
35 51 50.0 362 2 I37492 HLA-B alpha-chain  
36 51 50.0 362 2 I54505 lymphocyte antigen  
37 51 50.0 362 2 I56130 HLA-B\*5401 - human  
38 51 50.0 362 2 I36962 MHC class I protei  
39 51 50.0 362 2 I37475 HLA-B35 variant -  
40 51 50.0 362 2 I56149 lymphocyte antigen  
41 51 50.0 362 2 I59655 lymphocyte antigen  
42 51 50.0 362 2 I59651 lymphocyte antigen  
43 51 50.0 362 2 I61865 MHC HLA-B\*42, HLA-  
44 51 50.0 362 2 I61859 MHC HLA-B14 chain  
45 51 50.0 362 2 I61860 MHC HLA-B18 chain

#### ALIGNMENTS

##### RESULT 1

I38875

MHC class I antigen - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I38875

R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stef

Immunogenetics 42, 19-27, 1995

A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.

A:Reference number: I38860; MUID:95317819

A:Accession: I38875

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-137 <RES>

A:Cross-references: EMBL:U15639; NID:g930332; PIDN:AAA74046.1; PID:g930333

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 137;

Best Local Similarity 100.0%; Pred. No. 0.91;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRGY 20

DB 40 RESLRNLRGY 49

##### RESULT 2

I38876

MHC class I antigen - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I38876

R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stef

Immunogenetics 42, 19-27, 1995

A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.

A:Reference number: I38860; MUID:95317819

A:Accession: I38876

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-137 <RES>

A:Cross-references: EMBL:U15640; NID:g930334; PIDN:AAA74047.1; PID:g930335

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 137;

Best Local Similarity 100.0%; Pred. No. 0.91;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRGY 20

DB 40 RESLRNLRGY 49

```
RESULT 3
I38860
MHC class I antigen - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 23-Jul-1999
C:Accession: I38860
R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stefani
Immunogenetics 42, 19-27, 1995
A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.
A:Reference number: I38860; MUID:95317819
A:Accession: I38860
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-137 <RES>
A:Cross-references: EMBL:U14756; NID:g930328; PIDN:AA50171.1; PID:g930329
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNRLGY 20
Db 40 RESLNRLGY 49

RESULT 4
I38874
MHC class I antigen - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 23-Jul-1999
C:Accession: I38874
R:Garber, T.L.; Butler, L.M.; Trachtenberg, E.A.; Erlich, H.A.; Rickards, O.; De Stefani
Immunogenetics 42, 19-27, 1995
A:Title: HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles.
A:Reference number: I38860; MUID:95317819
A:Accession: I38874
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-137 <RES>
A:Cross-references: EMBL:U15638; NID:g930330; PIDN:AA74045.1; PID:g930331
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNRLGY 20
Db 40 RESLNRLGY 49

RESULT 5
I59188
MHC cell surface glycoprotein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999
C:Accession: I59188
R:Bronson, S.K.; Pei, J.; Taillon-Miller, P.; Chorney, M.J.; Geraghty, D.E.; Chaplin, D.
Proc. Natl. Acad. Sci. U.S.A. 88, 1676-1680, 1991
A:Title: Isolation and characterization of yeast artificial chromosome clones linking th
A:Reference number: I59188; MUID:91156671
A:Accession: I59188
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-181 <RES>
A:Cross-references: GB:M59841; NID:g187697; PIDN:AAA59623.1; PID:g187698
C:Genetics:
A:Gene: GDB:HLA-B
A:Cross-references: GDB:120048; OMIM:142830

Query Match 50.0%; Score 51; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNRLGY 20
Db 40 RESLNRLGY 49
```

```
A:Map position: 6p21.3-6p21.3
A:Introns: 90/1
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: glycoprotein

Query Match 50.0%; Score 51; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNRLGY 20
Db 74 RESLNRLGY 83

RESULT 6
HLH040
MHC class I histocompatibility antigen HLA-B40 alpha chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 02-Sep-1997
C:Accession: A02186
R:Lopez de Castro, J.A.; Bragado, R.; Strong, D.M.; Strominger, J.L.
Biochemistry 22, 3961-3969, 1983
A:Title: Primary structure of papain-solubilized human histocompatibility antigen HLA
A:Reference number: A02186; MUID:84000412
A:Accession: A02186
A:Molecule type: protein
A:Residues: 1-270 <LOP>
A:Comment: Comparison of a number of class I allelic sequences (-B40 vs. -B7 and -A2
y restricted to the alpha-1 and alpha-2 domains. The most conspicuous clustering of v
ossible alloantigenic determinants of these antigens.
C:Genetics:
A:Gene: GDB:HLA-B
A:Cross-references: GDB:120048; OMIM:142830
A:Map position: 6p21.3-6p21.3
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterodimer; membrane protein; transplantation
F;1-90/Domain: alpha-1 <EX1>
F;91-181/Domain: alpha-2 <EX2>
F;195-260/Domain: immunoglobulin homology <IM>
F;86/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;101-163,202-258/Disulfide bonds: #status predicted

Query Match 50.0%; Score 51; DB 1; Length 270;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNRLGY 20
Db 75 RESLNRLGY 84

RESULT 7
I68774
MHC HLA-B39 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 23-Jul-1999
C:Accession: I68774
R:Mueller, C.A.; Engler-Blum, G.; Gekeler, V.; Steiert, I.; Weiss, E.; Schmidt, H.
Immunogenetics 30, 200-207, 1989
A:Title: Genetic and serological heterogeneity of the supertypic HLA-B locus specific
A:Reference number: I54463; MUID:89379286
A:Accession: I68774
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-274 <RES>
A:Cross-references: GB:M29865; NID:g187676; PIDN:AAA36223.1; PID:g187677
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 1.8;
```

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRCY 20  
|||||

Db 75 RESLRNLRCY 84

## RESULT 8

S24439  
class I histocompatibility antigen HLA-B-3901 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 01-Nov-1996 #text\_change 07-Nov-1997  
C:Accession: S24439  
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Garbe  
Nature 357, 329-333, 1992  
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indicate  
A:Reference number: S24027; MUID:92269956  
A:Molecule type: mRNA  
A:Residues: 1-274 <WAT>  
C:Genetics:  
A:Gene: HLA-B-3901  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein  
F:1-274/Product: class I histocompatibility antigen HLA-B-3901 #status predicted <WAT>  
F:1-90/Domain: alpha-1 #status predicted <EX1>  
F:91-182/Domain: alpha-2 #status predicted <EX2>  
F:196-261/Domain: immunoglobulin homology <IM>  
F:86/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:101-164,203-259/disulfide bonds: #status predicted

Query Match 50.0%; Score 51; DB 2; Length 274;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRCY 20  
|||||

Db 75 RESLRNLRCY 84

## RESULT 9

I68701  
cell surface antigen - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
C:Accession: I68701  
R:Arnot, D.; Lillie, J.W.; Aufrey, C.; Kappes, D.; Strominger, J.L.  
Immunogenetics 20, 237-252, 1984  
A:Title: Inter-locus and intra-allelic polymorphisms of HLA class I antigen gene mRNA.  
A:Reference number: I54412; MUID:84287690  
A:Accession: I68701  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-300 <RES>  
A:Cross-references: GB:M27540; NID:g187733; PIDN:AAA59638.1; PID:g386890  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: surface antigen

Query Match 50.0%; Score 51; DB 2; Length 300;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRCY 20  
|||||

Db 37 RESLRNLRCY 46

## RESULT 10

I36956  
MHC class II chain - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999  
C:Accession: I36956  
R:Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.  
J. Immunol. 142, 3937-3950, 1989  
A:Title: Diversity and diversification of HLA-A,B,C alleles.  
A:Reference number: I36956; MUID:89235215  
A:Accession: I36956  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-308 <RES>  
A:Cross-references: GB:M24044; NID:g176812; PIDN:AAA35423.1; PID:g176813  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 308;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRCY 20  
|||||

Db 45 RESLRNLRCY 54

## RESULT 11

I68747  
MHC class I lymphocyte antigen - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
C:Accession: I68747  
R:Pohla, H.; Kuon, W.; Tabaczewski, P.; Doerner, C.; Weiss, E.H.  
Immunogenetics 29, 297-307, 1989  
A:Title: Allelic variation in HLA-B and HLA-C sequences and the evolution of the HLA-  
A:Reference number: I54457; MUID:89233295  
A:Accession: I68747  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-350 <RES>  
A:Cross-references: GB:M28204; NID:g576472; PIDN:AAA53257.1; PID:g576473  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 350;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLRCY 20  
|||||

Db 87 RESLRNLRCY 96

## RESULT 12

I54308  
MHC class I B71 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999  
C:Accession: I54308  
R:Rodriguez, S.G.; Johnson, A.H.; Hurley, C.K.  
Hum. Immunol. 37, 192-194, 1993  
A:Title: Molecular characterization of HLA-B71 from an African American individual.  
A:Reference number: I54308; MUID:94064392  
A:Accession: I54308  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-350 <RES>  
A:Cross-references: GB:L07950; NID:g307236; PIDN:AAA59683.1; PID:g307237  
C:Genetics:  
A:Gene: GDB:HLA-B  
A:Cross-references: GDB:120048; OMIM:142830  
A:Map position: 6p21.3-6p21.3  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 50.0%; Score 51; DB 2; Length 350;

Best Local Similarity 100.0%; Score 51; DB 2; Length 354;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLGY 20  
| | | | | | | | | |  
DB 99 RESLNRLGY 108

RESULT 13  
S24436  
Class I histocompatibility antigen HLA-B-3504 precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 18-Jul-1997 #text\_change 07-Nov-1997  
C:Accession: S24436  
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Garbe  
Nature 357, 329-333, 1992  
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indicated  
A:Reference number: S24027; MUID:92269956  
A:Accession: S24436  
A:Molecule type: mRNA  
A:Residues: 1-354 <WAT>  
C:Genetics:  
A:Gene: HLA-B-3504  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; transmembrane protein  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:17-354/Product: class I histocompatibility antigen HLA-B-3504 #status predicted <WAT>  
F:17-299/Domain: extracellular #status predicted <EXT>  
F:107-198/Domain: alpha-2 #status predicted <EX2>  
F:212-277/Domain: immunoglobulin homology <IMM>  
F:300-323/Domain: transmembrane #status predicted <TM>  
F:324-354/Domain: intracellular #status predicted <INT>  
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:117-180,219-275/Disulfide bonds: #status predicted

Query Match 50.0%; Score 51; DB 2; Length 354;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLGY 20  
| | | | | | | | | |  
DB 91 RESLNRLGY 100

RESULT 14  
S24437  
Class I histocompatibility antigen HLA-B-4802 precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 18-Jul-1997 #text\_change 07-Nov-1997  
C:Accession: S24437  
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Garbe  
Nature 357, 329-333, 1992  
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indicated  
A:Reference number: S24027; MUID:92269956  
A:Accession: S24437  
A:Molecule type: mRNA  
A:Residues: 1-354 <WAT>  
C:Genetics:  
A:Gene: HLA-B-4802  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; transmembrane protein  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:17-354/Product: class I histocompatibility antigen HLA-B-4802 #status predicted <WAT>  
F:17-299/Domain: extracellular #status predicted <EXT>  
F:107-198/Domain: alpha-2 #status predicted <EX2>  
F:212-277/Domain: immunoglobulin homology <IMM>  
F:300-323/Domain: transmembrane #status predicted <TM>  
F:324-354/Domain: intracellular #status predicted <INT>  
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:117-180,219-275/Disulfide bonds: #status predicted

Query Match 50.0%; Score 51; DB 2; Length 354;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLGY 20  
| | | | | | | | | |  
DB 91 RESLNRLGY 100

RESULT 15  
S24438  
Class I histocompatibility antigen HLA-B-4801 precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 18-Jul-1997 #text\_change 07-Nov-1997  
C:Accession: S24438  
R:Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford, E.L.; Levine, C.G.; Ga  
Nature 357, 329-333, 1992  
A:Title: New recombinant HLA-B alleles in a tribe of South American Amerindians indic  
A:Reference number: S24027; MUID:92269956  
A:Accession: S24438  
A:Molecule type: mRNA  
A:Residues: 1-354 <WAT>  
C:Genetics:  
A:Gene: HLA-B-4801  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; transmembrane protein  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:17-354/Product: class I histocompatibility antigen HLA-B-4801 #status predicted <WA  
F:17-299/Domain: extracellular #status predicted <EXT>  
F:107-198/Domain: alpha-2 #status predicted <EX2>  
F:212-277/Domain: immunoglobulin homology <IMM>  
F:300-323/Domain: transmembrane #status predicted <TM>  
F:324-354/Domain: intracellular #status predicted <INT>  
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:117-180,219-275/Disulfide bonds: #status predicted

Query Match 50.0%; Score 51; DB 2; Length 354;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLGY 20  
| | | | | | | | | |  
DB 91 RESLNRLGY 100

Search completed: February 7, 2000, 18:04:41  
Job time: 22207 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:26:01 ; Search time 68.63 Seconds  
(without alignments)  
8.703 Million cell updates/sec

Title: US-08-653-294-35

Perfect score: 102

Sequence: 1 YCRLNLSERRESLRLRGY 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	50.0	270	1	P01890 homo sapien
2	51	50.0	354	1	P30470 homo sapien
3	51	50.0	362	1	P16210 pan troglod
4	51	50.0	362	1	P01889 homo sapien
5	51	50.0	362	1	P13751 pan troglod
6	51	50.0	362	1	P30450 homo sapien
7	51	50.0	362	1	P30462 homo sapien
8	51	50.0	362	1	P30453 homo sapien
9	51	50.0	362	1	P30464 homo sapien
10	51	50.0	362	1	P30465 homo sapien
11	51	50.0	362	1	P30513 homo sapien
12	51	50.0	362	1	P30456 homo sapien
13	51	50.0	362	1	P30685 homo sapien
14	51	50.0	362	1	P30468 homo sapien
15	51	50.0	362	1	P30469 homo sapien
16	51	50.0	362	1	P30471 homo sapien
17	51	50.0	362	1	P30472 homo sapien
18	51	50.0	362	1	P30473 homo sapien
19	51	50.0	362	1	P30474 homo sapien
20	51	50.0	362	1	P30475 homo sapien
21	51	50.0	362	1	P30476 homo sapien
22	51	50.0	362	1	Q04826 homo sapien
23	51	50.0	362	1	P30477 homo sapien
24	51	50.0	362	1	P30478 homo sapien
25	51	50.0	362	1	P30479 homo sapien
26	51	50.0	362	1	P30480 homo sapien
27	51	50.0	362	1	P30483 homo sapien
28	51	50.0	362	1	P30486 homo sapien
29	51	50.0	362	1	P30488 homo sapien
30	51	50.0	362	1	P30492 homo sapien
31	51	50.0	362	1	P30493 homo sapien
32	51	50.0	362	1	P30494 homo sapien
33	51	50.0	362	1	P30495 homo sapien
34	51	50.0	362	1	P30496 homo sapien

```

35 51 50.0 362 1 1B63_HUMAN
36 47.5 46.6 605 1 GLCA_SOYBN
37 46 45.1 346 1 HRRT_PSESY
38 45.5 44.6 366 1 1C01_HUMAN
39 44.5 43.6 639 1 GLCK_SOYBN
40 44 43.1 342 1 1CXX_HUMAN
41 44 43.1 362 1 1B44_HUMAN
42 44 43.1 366 1 1C01_PANTR
43 44 43.1 366 1 1C02_HUMAN
44 44 43.1 366 1 1C05_HUMAN
45 44 43.1 366 1 1C06_HUMAN

```

```

P30498 homo sapien
P13916 glycine max
Q05395 pseudomonas
P30499 homo sapien
P11827 glycine max
P10321 homo sapien
P30484 homo sapien
P30686 pan troglod
P30500 homo sapien
P04322 homo sapien
P30503 homo sapien

```

#### ALIGNMENTS

```

RESULT 1
ID 1B33_HUMAN STANDARD; PRT: 270 AA.
AC P01890;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-60(B-40) B*4001 ALPHA CHAIN
DE (FRAGMENT).
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE.
RX MEDLINE: 84000412.
RA LOPEZ DE CASTRO J.A., BRAGADO R., STRONG D.M., STROMINGER J.L.;
RT "Primary structure of papain-solubilized human histocompatibility
RT antigen HLA-B40 (-Bw60). An outline of alloantigenic determinants.";
RL Biochemistry 22:3961-3969(1983).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC PIR: A02186; HLHU40.
DR HSSP: P30460; LAGB.
DR MIM: 142830; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
DR PFAM: PF00129; MHC_I; 1.
DR MHC I; Glycoprotein.
KW DOMAIN 1 90
FT DOMAIN 91 181
FT DOMAIN 182 >270
FT CARBOHYD 86 86
FT DISULFID 101 163
FT DISULFID 202 238
FT NON_TER 270 270
SQ SEQUENCE 270 AA; 31205 MW; BFE44EFF CRC32;

```

```

EXTRACELLULAR ALPHA-1.
EXTRACELLULAR ALPHA-2.
EXTRACELLULAR ALPHA-3.

```

Query Match 50.0%; Score 51; DB 1; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 11 RESLRNLRGY 20
  |||||
Db 75 RESLRNLRGY 84

```

```

RESULT. 2
ID 1B24_HUMAN STANDARD; PRT: 354 AA.
AC P30470;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)

```

[illegible]



RT	"Diversity of class I HLA molecules: functional and evolutionary
RT	interactions with T cells.";
RL	Cold Spring Harb. Symp. Quant. Biol. 54:529-543(1989).
RN	[3]
RN	SEQUENCE FROM N.A.
RX	MEDLINE: 85287366.
RA	SOOD A.K., PAN J., BIRO P.A., PEREIRA D., SRIVASTAVA R., REDDY V.B.,
RA	DUCEMAN B.W., WEISSMAN S.M.;
RL	"Structure and polymorphism of class I MHC antigen mRNA.";
RL	Immunogenetics 22:101-121(1985).
RN	[4]
RN	SEQUENCE FROM N.A.
RP	EILEXSON M.E., ZHANG L., HILDEBRAND W.H.;
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN	[5]
RP	SEQUENCE OF 25-295.
RX	MEDLINE: 80088278.
RA	ORR H.T., LOPEZ DE CASTRO J.A., LANCET D., STROMINGER J.L.;
RT	"Complete amino acid sequence of a papain-solubilized human
RT	histocompatibility antigen, HLA-B*7. 2. Sequence determination and
RT	search for homologues.";
RL	Biochemistry 18:5711-5720(1979).
CC	-1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC	THE IMMUNE SYSTEM.
CC	-1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC	MICROGLOBULIN).

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to licensee@isb-sib.ch).
CC	-----
CC	EMBL; M32317; AAA36230.1; ..
DR	EMBL; M16102; AAA59622.1; ALT_SEQ.
DR	EMBL; U29057; AAA91229.1; ..
DR	PIR; A02185; HLHUB7.
DR	PIR; B35997; B35997.
DR	HSSP; P30460; IAGB.
DR	MIM; 142830; ..
DR	PROSITE; PS00290; IG_MHC; 1.
DR	PFAM; PF00047; ig; 1. 1.
DR	FRAM; FR00129; MHC-I; 1.
DR	MHC I; Transmembrane; Glycoprotein; Signal.
FT	SIGNAL 1 24
FT	CHAIN 25 362
FT	
FT	DOMAIN 25 114
FT	DOMAIN 115 206
FT	DOMAIN 207 298
FT	DOMAIN 299 309
FT	TRANSMEM 310 333
FT	DOMAIN 334 362
FT	CARBOHYD 110 110
FT	DISULFID 125 188
FT	DISULFID 227 283
FT	CONFLICT 15 18
FT	CONFLICT 266 266
FT	CONFLICT 268 268
FT	CONFLICT 297 297
FT	CONFLICT 314 315
FT	SEQUENCE 362 AA;
SQ	SEQUENCE 40406 MW; 87B2ED84 CRC32;
	AALA -> GPW (IN REF. 3).
	O -> E (IN REF. 5).
	W -> S (IN REF. 3).
	R -> G (IN REF. 3).
	GL -> RP (IN REF. 3).

```

Query Match          50.0%; Score 51; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11  RESLRNLRGY 20
          |||||
Db       99  RESLRNLRGY 108

```

RESULT	5
ID	1B02_PANTR
ID	1B02_PANTR STANDARD; PRT; 362 AA.
AC	P13751;
DT	01-JAN-1990 (Rel. 13, Created)
DT	01-JAN-1990 (Rel. 13, Last sequence update)
DT	01-APR-1993 (Rel. 25, Last annotation update)
DE	CHIA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-2 ALPHA CHAIN PRECURSOR.
OS	Pan troglodytes (Chimpanzee).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC	Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN	[1]
RN	SEQUENCE FROM N.A.
RX	MEDLINE: 89030641.
RA	MAYER W.E., JONKER M., KLEIN D., IVANYI P., VAN SEVENTER G.,
RA	KLEIN J.;
RT	"Nucleotide sequences of chimpanzee MHC class I alleles: evidence for
RT	trans-species mode of evolution.";
RL	EMBO J. 7:2765-2774(1988).
RN	[2]
RN	REVISIONS.
RP	
RA	MAYER W.;
RA	Submitted (FEB-1989) to the EMBL/GenBank/DDBJ databases..
CC	-1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC	THE IMMUNE SYSTEM.
CC	
CC	-1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC	MICROGLOBULIN).
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	
DR	EMBL: X13116; CAA31508.1; ..
DR	PIR: S03538; S03538.
DR	HSP: P30685; IAIIN.
DR	PROSITE: PS00290; IG_MHC; 1.
DR	PFAM: PF00047; Ig; 1.
DR	PFAM: PF00129; MHC_I; 1.
KW	MHC I; Transmembrane; Glycoprotein; Signal.
FT	SIGNAL
FT	1 24
FT	CHAIN 25 362
FT	
FT	DOMAIN 25 114
FT	EXTRACELLULAR ALPHA-1.
FT	DOMAIN 115 206
FT	EXTRACELLULAR ALPHA-2.
FT	DOMAIN 207 298
FT	EXTRACELLULAR ALPHA-3.
FT	DOMAIN 299 308
FT	CONNECTING PEPTIDE.
FT	DOMAIN 309 332
FT	CYTOPLASMIC TAIL.
FT	DOMAIN 333 362
FT	BY SIMILARITY.
FT	DISULFID 125 188
FT	BY SIMILARITY.
FT	DISULFID 227 283
FT	BY SIMILARITY.
FT	CARBOHYD 110 110
FT	SEQUENCE 362 AA; 40488 MW; 4BF65A6C CRC32;
SQ	

```

Query Match          50.0%; Score 51; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. NO. 0.82;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 RESRLNLRGY 20
         |||||
Db       99 RESRLNLRGY 108

RESULT      6
1B04_HUMAN
ID 1B04_HUMAN STANDARD; PRT; 362 AA.
AC P30460;

```

DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-8 B\*0801 ALPHA CHAIN  
DE PRECURSOR.  
GN HLA-B OR HLAB.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 89235215.  
RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
RT "Diversity and diversification of HLA-A,B,C alleles."; J. Immunol. 142:3937-3950(1989).

RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 25-300.  
RX REID S.W., MCADAM S., SMITH K.J., KLERNERMAN P., O'CALLAGHAN C.A.,  
RA HARLOS K., JAKOBSEN B.K., McMICHAEL A.J., BELL J.I., STUART D.I.,  
RJ JONES E.Y.;  
RT "Antagonist HIV-1 Gag peptides induce structural changes in HLA B8." J. Exp. Med. 184:2279-2286(1996).

-! FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
CC -! SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

EMBL: M24040; AAA52662.1;  
PDB: 1AGB; 16-JUN-97.  
PDB: 1AGC; 16-JUN-97.  
PDB: 1AGE; 16-JUN-97.  
PDB: 1AGF; 16-JUN-97.  
MIM: 142830;  
PROSITE: PS00290; IG\_MHC; 1.  
PFAM: PF00047; ig\_1.  
MHC\_I; Transmembrane; Glycoprotein; Signal.  
CHAIN 1 24

HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
B-8 B\*0801 ALPHA CHAIN.  
EXTRACELLULAR ALPHA-1.  
EXTRACELLULAR ALPHA-2.  
CONNECTING PEPTIDE.  
CYTOPLASMIC TAIL.  
BY SIMILARITY.  
DISULFID 125 188 BY SIMILARITY.  
SEQUENCE 362 AA: 40331 MW; 146789EB CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 0.82;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dt 11 RESRLNRGY 20  
Db 99 RESRLNRGY 108

RESULT 7  
ID 1B07\_HUMAN STANDARD; PRT; 362 AA.  
AC F30462;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE 01-APR-1993 (Rel. 25, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-8 B\*0801 ALPHA CHAIN  
DE PRECURSOR.  
GN HLA-B OR HLAB.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 89235215.  
RA FARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
RT "Diversity and diversification of HLA-A,B,C alleles."; J. Immunol. 142:3937-3950(1989).

RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTRÖMS) OF 25-300.  
RX REID S.W., MCADAM S., SMITH K.J., KLENERMAN P., O'CALLAGHAN C.A.,  
RA HARLOS K., JAKOBSEN B.K., McMICHAEI A.J., BELL J.I., STUART D.I.,  
RJ JONES E.Y.;  
RT "Antagonist HIV-1 Gag peptides induce structural changes in HLA B8." J. Exp. Med. 184:2279-2286(1996).

-! FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
CC -! SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

EMBL: M24040; AAA52662.1;  
PDB: 1AGB; 16-JUN-97.  
PDB: 1AGC; 16-JUN-97.  
PDB: 1AGE; 16-JUN-97.  
PDB: 1AGF; 16-JUN-97.  
MIM: 142830;  
PROSITE: PS00290; IG\_MHC; 1.  
PFAM: PF00047; ig\_1.  
MHC\_I; Transmembrane; Glycoprotein; Signal.  
CHAIN 1 24

HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
B-8 B\*0801 ALPHA CHAIN.  
EXTRACELLULAR ALPHA-1.  
EXTRACELLULAR ALPHA-2.  
CONNECTING PEPTIDE.  
CYTOPLASMIC TAIL.  
BY SIMILARITY.  
DISULFID 125 188 BY SIMILARITY.  
SEQUENCE 362 AA: 40331 MW; 146789EB CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 0.82;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dt 11 RESRLNRGY 20  
Db 99 RESRLNRGY 108

RESULT 8  
ID 1B08\_HUMAN STANDARD; PRT; 362 AA.  
AC F30463;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE 01-APR-1993 (Rel. 25, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-65(B-14) B\*1402 ALPHA CHAIN  
DE PRECURSOR.  
GN HLA-B OR HLAB.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

DT	01-APR-1993	(Rel. 25, Created)
DT	01-APR-1993	(Rel. 25, Last sequence update)
DT	15-JUL-1998	(Rel. 36, Last annotation update)
DE	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-8 B*0801 ALPHA CHAIN	
DE	PRECURSOR.	
GN	HLA-B OR HLAB.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;	
OC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 89235215.	
RA	PARRHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;	
RT	"Diversity and diversification of HLA-A,B,C alleles.";	
RL	J. Immunol. 142:3937-3950(1989).	
CC	-!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.	
CC	-!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).	
DR	EMBL; M24040; AAA52662.1; -	
DR	HSSP; P30460; 1AGB.	
DR	MIM; 142830; -	
DR	PROSITE; PS00290; IG_MHC; 1.	
DR	PFAM; PF00047; ig_1.	
DR	MHC I; Transmembrane; Glycoprotein; Signal.	
KW	SIGNAL 1 24	
FT	CHAIN 25 362	
FT	DOMAIN 25 114	
FT	DOMAIN 115 206	
FT	DOMAIN 207 298	
FT	DOMAIN 299 309	
FT	TRANSMEM 310 333	
FT	DOMAIN 334 362	
FT	CARBOHYD 110 110	
FT	DISULFID 125 188	
FT	DISULFID 227 283	
SQ	SEQUENCE 362 AA; 40331 MW; 1467B9EB CRC32;	
Query Match 50.0%; Score 51; DB 1; Length 362;		
Best Local Similarity 100.0%; Pred. No. 0.82;		
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	11 RESLRNLRCY 20	
DB	99 RESLRNLRCY 108	
RESULT 8		
ID	1B07_HUMAN	
AC	P30463;	
DT	01-APR-1993 (Rel. 25, Created)	
DT	01-APR-1993 (Rel. 25, Last sequence update)	
DE	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-14 B*1401 ALPHA CHAIN	
DE	PRECURSOR.	
GN	HLA-B OR HLAB.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 89235215.	
RA	PARRHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;	
RT	"Diversity and diversification of HLA-A,B,C alleles.";	
RL	J. Immunol. 142:3937-3950(1989).	
CC	-!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.	
CC	-!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).	
DR	EMBL; M24040; AAA52662.1; -	
DR	HSSP; P30460; 1AGB.	
DR	MIM; 142830; -	
DR	PROSITE; PS00290; IG_MHC; 1.	
DR	PFAM; PF00047; ig_1.	
DR	MHC I; Transmembrane; Glycoprotein; Signal.	
KW	SIGNAL 1 24	
FT	CHAIN 25 362	
FT	DOMAIN 25 114	
FT	DOMAIN 115 206	
FT	DOMAIN 207 298	
FT	DOMAIN 299 309	
FT	TRANSMEM 310 333	
FT	DOMAIN 334 362	
FT	CARBOHYD 110 110	
FT	DISULFID 125 188	
FT	DISULFID 227 283	
SQ	SEQUENCE 362 AA; 40331 MW; 1467B9EB CRC32;	
Query Match 50.0%; Score 51; DB 1; Length 362;		
Best Local Similarity 100.0%; Pred. No. 0.82;		
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	11 RESLRNLRCY 20	
DB	99 RESLRNLRCY 108	
RESULT 7		
ID	1B07_HUMAN	
AC	P30463;	
DT	01-APR-1993 (Rel. 25, Created)	
DT	01-APR-1993 (Rel. 25, Last sequence update)	
DE	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-65(B-14) B*1402 ALPHA CHAIN	
DE	PRECURSOR.	
GN	HLA-B OR HLAB.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 89235215.	
RA	PARRHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;	
RT	"Diversity and diversification of HLA-A,B,C alleles.";	
RL	J. Immunol. 142:3937-3950(1989).	
CC	-!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.	
CC	-!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).	
DR	EMBL; M24040; AAA52662.1; -	
DR	HSSP; P30460; 1AGB.	
DR	MIM; 142830; -	
DR	PROSITE; PS00290; IG_MHC; 1.	
DR	PFAM; PF00047; ig_1.	
DR	MHC I; Transmembrane; Glycoprotein; Signal.	
KW	SIGNAL 1 24	
FT	CHAIN 25 362	
FT	DOMAIN 25 114	
FT	DOMAIN 115 206	
FT	DOMAIN 207 298	
FT	DOMAIN 299 309	
FT	TRANSMEM 310 333	
FT	DOMAIN 334 362	
FT	CARBOHYD 110 110	
FT	DISULFID 125 188	
FT	DISULFID 227 283	
SQ	SEQUENCE 362 AA; 40331 MW; 1467B9EB CRC32;	
Query Match 50.0%; Score 51; DB 1; Length 362;		
Best Local Similarity 100.0%; Pred. No. 0.82;		
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	11 RESLRNLRCY 20	
DB	99 RESLRNLRCY 108	

OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE; 89235215.  
 RX PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
 RA "Diversity and diversification of HLA-A,B,C alleles."  
 RT J. Immunol. 142:3937-3950(1989).  
 RL  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M24032; AAA59664.1; -  
 DR HSP: P30460; 1AGB.  
 DR MM: 142830; -  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 25 114 BW-65(B-14) B\*1402 ALPHA CHAIN.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 299 309 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 310 333 CONNECTING PEPTIDE.  
 FT DOMAIN 334 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40342 MW; BE68AC9E CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNLRGY 20  
 Db 99 RESLNLRGY 108

RESULT 9  
 1B10\_HUMAN  
 ID 1B10\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30464;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-75(B-15) B\*1502 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92196792.  
 RA "The HLA-B\*75 subtype of B15: molecular characterization and  
 RT comparison with crossreacting antigens."  
 RT Tissue Antigens 38:186-190(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC TISSUE-BLOOD;  
 RX MEDLINE; 96369309.  
 RA LIN L., TOKUNAGA K., TANAKA H., NAKAJIMA F., IMANISHI T.,  
 RA KASHIWASE K., BANNAI M., MIZUNO S., AKAZA T., TADOKORO K.,  
 RA SHIBATA Y., JUJI T.;  
 RT "Further molecular diversity in the HLA-B15 group."  
 RL Tissue Antigens 47:265-274(1996).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M75138; AAA59630.1; -  
 DR HSP: P30460; 1AGB.  
 DR MM: 142830; -  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 25 114 BW-75(B-15) B\*1502 ALPHA CHAIN.  
 FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 299 309 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 310 333 CONNECTING PEPTIDE.  
 FT DOMAIN 334 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40338 MW; 8CF9BCD0 CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNLRGY 20  
 Db 99 RESLNLRGY 108

RESULT 10  
 1B11\_HUMAN  
 ID 1B11\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30465;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-72(BW-70) B\*1503 ALPHA  
 DE CHAIN PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93056508.  
 RA MADRIGAL J.A., BELICH M.P., HILDEBRAND W.H., BENJAMIN R.J.,  
 RA LITTLE A.-M., ZEMMOUR J., ENNIS P.D., WARD F.E., PETZL-ERLER M.L.,  
 RA MARTELL R.W., DU TOIT E.D., PARHAM P.;  
 RT "Distinctive HLA-A,B antigens of black populations formed by  
 RT interallelic conversion."

RL J. Immunol. 149:3411-3415(1992).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CG -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X61709; CAA43878.1; -;  
 DR PIR; S16789; S16789.  
 DR HSP; P30685; IAIN.  
 DR MIM; 142830; -;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT BW-72 (BW-70) B\*1503 ALPHA CHAIN.  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 309  
 FT TRANSMEM 310 333  
 FT DOMAIN 334 362  
 FT CARBOHYD 110 110  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 SQ SEQUENCE 362 AA; 40387 MW; 99D70546 CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 11 RESLRNLRGY 20  
 Db 99 RESLRNLRGY 108  
 |||||

RESULT 11  
 1B12\_HUMAN  
 ID 1B12\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30513;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-62 B\*1504 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92269956.  
 RA WATKINS D.I., MCADAM S.N., LIU X., STANG C.R., MILFORD E.L.,  
 RA LEVINE C.G., GABER T.L., DOGON A.L., LORD C.I., GHIM S.H.,  
 RA TROUP G.M., HUGHES A.L., LETVIN N.L.;  
 RT "New recombinant HLA-B alleles in a tribe of South American  
 RL Amerindians indicate rapid evolution of MHC class I loci.";  
 RL Nature 357:329-333(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BLOOD;  
 RA RAMOS M., BARBER D.F., LAYRISSE Z., DE CASTRO J.A.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M84382; AAA59632.1; -;  
 DR EMBL; U70528; AAB16918.1; -;  
 DR PIR; S24433; S24433.  
 DR HSP; P30685; IAIN.  
 DR MIM; 142830; -;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT B-62 B\*1504 ALPHA CHAIN.  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 309  
 FT TRANSMEM 310 333  
 FT DOMAIN 334 362  
 FT CARBOHYD 110 110  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 SQ SEQUENCE 362 AA; 40406 MW; E80FC24C CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 11 RESLRNLRGY 20  
 Db 99 RESLRNLRGY 108  
 |||||

RESULT 12  
 1B13\_HUMAN  
 ID 1B13\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30466;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-18 B\*1801 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89235215.  
 RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;  
 RT "Diversity and diversification of HLA-A,B,C alleles";  
 RL J. Immunol. 142:3937-3950(1989).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M24039; AAA59662.1; -  
 DR HSSP; P30460; IAGB.  
 DR MIM; 142830;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 309  
 FT TRANSMEM 310 333  
 FT DOMAIN 334 362  
 FT CARBOHYD 110 110  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 SQ SEQUENCE 362 AA; 40275 MW; 038EC3FC CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 |||||  
 DB 99 RESLNRLRGY 108

RESULT 13  
 1B21\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30685;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3501 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89339610.  
 RA COBA T., HAYASHI H., KARAKI S., TANABE M., KANO K., TAKIGUCHI M.;  
 RT "The structure of HLA-B\*35 suggests that it is derived from HLA-B\*58  
 RT by two genetic mechanisms.";  
 RL Immunogenetics 30:76-80(1989).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 25-300.  
 RX MEDLINE; 96209671.  
 RA SMITH K.J., REID S.W., STUART D.I., MCMICHAEL A.J., JONES E.Y.,  
 RA BELL J.I.;  
 RT "An altered position of the alpha 2 helix of MHC class I is revealed  
 RT by the crystal structure of HLA-B\*3501.";  
 RL Immunity 4:203-214(1996).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).  
 RA MENSSSEN R., ORTH P., ZIEGLER A., SAENGER W.;  
 RL Submitted (APR-1998) to the PDB data bank.  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M28115; AAA59617.1;  
 DR EMBL; M28109; AAA59617.1; JOINED.  
 DR EMBL; M28110; AAA59617.1; JOINED.  
 DR EMBL; M28111; AAA59617.1; JOINED.  
 DR EMBL; M28112; AAA59617.1; JOINED.  
 DR EMBL; M28113; AAA59617.1; JOINED.  
 DR EMBL; M28114; AAA59617.1; JOINED.  
 DR PIR; A45880; A45880.  
 DR PDB; 1A1N; 08-APR-98.  
 DR PDB; 1A9B; 18-NOV-98.  
 DR PDB; 1A9E; 18-NOV-98.  
 DR MIM; 142830;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 308  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362  
 FT CARBOHYD 110 110  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 SQ SEQUENCE 362 AA; 40455 MW; AECIC675 CRC32;

HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 B-35 B\*3501 ALPHA CHAIN.  
 EXTRACELLULAR ALPHA-1.  
 EXTRACELLULAR ALPHA-2.  
 EXTRACELLULAR ALPHA-3.  
 CONNECTING PEPTIDE.

FT CYTOPLASMIC TAIL.  
 FT BY SIMILARITY.

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLNRLRGY 20  
 |||||  
 DB 99 RESLNRLRGY 108

RESULT 14  
 1B22\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30458;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3502 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91365651.  
 RA CHERNOFF L.P., HERRERA M., FAINBOIM L., SATZ M.L.;  
 RT "Completed nucleotide sequence of a genomic clone encoding HLA-B\*35  
 RT isolated from a Caucasian individual of Hispanic origin.  
 RT Identification of a new variant of HLA-B\*35.";  
 RL Hum. Immunol. 31:153-158(1991).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M63454; AAA59682.1; -;  
 DR HSSP; P30685; IAIN.  
 DR MIM; 142830; -;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT DOMAIN 25 114 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 115 206 B-35 B\*3503 ALPHA CHAIN.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.  
 FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
 FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110 BY SIMILARITY.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40564 MW; D7B5C2C9 CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNLRGY 20  
 Db 99 RESLNLRGY 108

RESULT 15  
 1B23\_HUMAN  
 ID 1B23\_HUMAN STANDARD; PRT; 362 AA.  
 AC P30469;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3503 ALPHA CHAIN  
 DE PRECURSOR.  
 GN HLA-B OR HLAB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92176661.  
 RA ZENMOUR J., LITTLE A.M., SCHENDEL D.J., PARHAM P.;  
 RT "The HLA-A,B 'negative' mutant cell line C1R expresses a novel  
 RT HLA-B\*35 allele, which also has a point mutation in the translation  
 RT initiation codon."  
 RL J. Immunol. 148:1941-1948(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BLOOD;  
 RX MEDLINE; 95279930.  
 RA BECK Y., SATZ L., TAKAMIYA Y., NAKAYAMA S., LING L., ISHIKAWA Y.,  
 RA NAGAO T., UCHIDA H., TOKUNAGA K., MULLER C., JUJI T., TAKIGUCHI M.;  
 RT "Polymorphism of human minor histocompatibility antigens: T cell  
 RT recognition of human minor histocompatibility peptides presented by  
 RT HLA-B\*35 subtype molecules."  
 RL J. Exp. Med. 181:2037-2048(1995).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.

CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M81798; AAA59684.1; -;  
 DR EMBL; D50299; BAA08828.1; -;  
 DR HSSP; P30685; IAIN.  
 DR MIM; 142830; -;  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; Ig; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT DOMAIN 25 114 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 115 206 B-35 B\*3503 ALPHA CHAIN.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-2.  
 FT TRANSMEM 309 332 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 333 362 CONNECTING PEPTIDE.  
 FT CARBOHYD 110 110 CYTOPLASMIC TAIL.  
 FT DISULFID 125 188 BY SIMILARITY.  
 FT DISULFID 227 283 BY SIMILARITY.  
 SQ SEQUENCE 362 AA; 40515 MW; 4222D30A CRC32;

Query Match 50.0%; Score 51; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RESLNLRGY 20  
 Db 99 RESLNLRGY 108

Search completed: February 8, 2000, 01:26:02  
 Job time: 1562 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	51	50.0	81	7	O19523	homo sapien
2	51	50.0	81	7	O19525	homo sapien
3	51	50.0	81	7	O19527	homo sapien
4	51	50.0	81	7	O19529	homo sapien
5	51	50.0	81	7	O19531	homo sapien
6	51	50.0	81	7	O19533	homo sapien
7	51	50.0	81	7	O19535	homo sapien
8	51	50.0	81	7	O19537	homo sapien
9	51	50.0	83	7	O19539	homo sapien
10	51	50.0	83	7	O19541	homo sapien
11	51	50.0	83	7	O19543	homo sapien
12	51	50.0	89	4	O9Y452	homo sapien
13	51	50.0	89	7	P79615	homo sapien
14	51	50.0	89	7	P79620	homo sapien
15	51	50.0	89	7	O19548	homo sapien
16	51	50.0	89	7	O19549	homo sapien
17	51	50.0	89	7	O19550	homo sapien
18	51	50.0	89	7	O19551	homo sapien
19	51	50.0	89	7	O19640	homo sapien
20	51	50.0	89	7	O77967	homo sapien

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014771; AAB67809.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
FT NON\_TER 81 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 50.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20  
DB 66 RESLRNLCY 75

RESULT 3  
O19527 PRELIMINARY; PRT; 81 AA.  
AC O19527;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014773; AAB67811.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
FT NON\_TER 81 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 50.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20  
DB 66 RESLRNLCY 75

RESULT 4  
O19529 PRELIMINARY; PRT; 81 AA.  
AC O19529;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.

RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014775; AAB67813.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
FT NON\_TER 81 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 50.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20  
DB 66 RESLRNLCY 75

RESULT 5  
O19531 PRELIMINARY; PRT; 81 AA.  
AC O19531;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014777; AAB67815.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1 1  
FT NON\_TER 81 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match 50.0%; Score 51; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20  
DB 66 RESLRNLCY 75

RESULT 6  
O19533 PRELIMINARY; PRT; 81 AA.  
AC O19533;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014779; AAB67817.1; -  
DR PFAM; PF00129; MHC\_I; 1.



KW MHC.  
FT NON\_TER 1  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

Query Match  
Best Local Similarity 50.0%; Score 51; DB 7; Length 81;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match  
Best Local Similarity 50.0%; Score 51; DB 7; Length 81;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRLRGY 20  
DB 66 RESLRLRGY 75

QY 11 RESLRLRGY 20  
DB 66 RESLRLRGY 75

RESULT 7  
O19535  
ID O19535 PRELIMINARY; PRT; 81 AA.

RESULT 9  
O19539  
ID O19539 PRELIMINARY; PRT; 83 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014781; AAB67819.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014785; AAB67823.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 83  
SQ SEQUENCE 83 AA; 9731 MW; 24B8D666 CRC32;

QY 11 RESLRLRGY 20  
DB 66 RESLRLRGY 75

QY 11 RESLRLRGY 20  
DB 66 RESLRLRGY 77

RESULT 8  
O19537  
ID O19537 PRELIMINARY; PRT; 81 AA.

RESULT 10  
O19541  
ID O19541 PRELIMINARY; PRT; 83 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014783; AAB67821.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014787; AAB67825.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 83  
SQ SEQUENCE 83 AA; 9731 MW; 24B8D666 CRC32;

Query Match  
Best Local Similarity 50.0%; Score 51; DB 7; Length 81;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match  
Best Local Similarity 50.0%; Score 51; DB 7; Length 83;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRLRGY 20  
DB 66 RESLRLRGY 75

QY 11 RESLRLRGY 20  
DB 66 RESLRLRGY 77

RESULT 8  
O19537  
ID O19537 PRELIMINARY; PRT; 81 AA.

RESULT 10  
O19541  
ID O19541 PRELIMINARY; PRT; 83 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014783; AAB67821.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 81  
SQ SEQUENCE 81 AA; 9405 MW; 073087CE CRC32;

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).  
GN HLA-B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,  
RA RUNGROUNG E., BEJCHANDRA S.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014787; AAB67825.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.  
FT NON\_TER 1  
FT NON\_TER 83  
SQ SEQUENCE 83 AA; 9731 MW; 24B8D666 CRC32;

Query Match  
Best Local Similarity 50.0%; Score 51; DB 7; Length 83;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match  
Best Local Similarity 50.0%; Score 51; DB 7; Length 83;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 11 RESLRNLCY 20
    |||||
Db 68 RESLRNLCY 77

RESULT 11
O19543
ID O19543 PRELIMINARY; PRT; 83 AA.
AC O19543;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MHC CLASS I ANTIGEN HLA-B (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHANDANAYINGYONG D., SIRIKONG M., LUANGTRAKOOL K., SRINAK D.,
RA RUNGRONG E., BEJCHANDRA S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF014789; AAB67827.1; -.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 83 83
SQ SEQUENCE 83 AA; 9731 MW; 24B8D666 CRC32;

Query Match 50.0%; Score 51; DB 7; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20
    |||||
Db 68 RESLRNLCY 77

RESULT 12
O9Y452
ID O9Y452 PRELIMINARY; PRT; 89 AA.
AC O9Y452;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HUMAN LEUCOCYTE ANTIGEN B (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CARTER V., DUNN P.P.;
RT "Identification of a novel HLA-B*07 allele.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237594; CAB40714.1; -.
FT NON_TER 1 1
FT NON_TER 89 89
SQ SEQUENCE 89 AA; 10609 MW; 659B735B CRC32;

Query Match 50.0%; Score 51; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20
    |||||
Db 74 RESLRNLCY 83

RESULT 13
P79615
ID P79615 PRELIMINARY; PRT; 89 AA.
AC P79615;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MHC CLASS I HLA-B (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHOPEK M., CAO K., ZHANG G.H.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88249; AAB48493.1; -.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 89 89
SQ SEQUENCE 89 AA; 10489 MW; A7D3DF93 CRC32;

Query Match 50.0%; Score 51; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20
    |||||
Db 74 RESLRNLCY 83

RESULT 14
P79620
ID P79620 PRELIMINARY; PRT; 89 AA.
AC P79620;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MHC CLASS I HLA-B (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHOPEK M., CAO K., ZHANG G.H.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88254; AAB48498.1; -.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 89 89
SQ SEQUENCE 89 AA; 10501 MW; FA3EFDCD CRC32;

Query Match 50.0%; Score 51; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RESLRNLCY 20
    |||||
Db 74 RESLRNLCY 83

RESULT 15
O19548
ID O19548 PRELIMINARY; PRT; 89 AA.
AC O19548;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MHC CLASS I ANTIGEN (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CAO K., BURDET L., ZHANG H., FERNANDEZ-VINA M.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF016299; AAB69444.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 FT NON\_TER 89  
 SQ SEQUENCE 89 AA; 10550 MW; 11452E40 CRC32;

Query Match 50.0%; Score 51; DB 7; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 RESLRNLRCY 20  
 DB 74 RESLRNLRCY 83

Search completed: February 8, 2000, 19:16:17  
 Job time: 21504 sec

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-35 to: GenEmbl.\* out\_format : pfs  
 Date: Feb 8, 2000 10:25 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
 -MODEL=frame+ p2n.model -DEV=xip  
 -O=/cgn1/1/USPTO.spool/US08653294/runat\_04022000\_160701\_15779/app\_query.fasta.2  
 -DB=GenEmbl -QFMT=fastap -SUFFIX=irge -GAPOP=12.000 -GAPEXT=4.000  
 -MINMATCH=0.100 -LOOPEXT=0.000 -XGAPEXT=0.500 -XGAPOP=6.000  
 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000  
 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000  
 -DELEX=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi  
 -LIST=45 -DOALIGN=200 -THR SCORE=1000000 -ALIGN=15 -MODE=LOCAL  
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-08-653-294-35

Query length: 20

Database: GenEmbl.\*

Database sequences: 821193

Database length: -1518192014

Search time (sec): 10176.920000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_pr2:ATAC007134	+ 52.00	95.99	2.5e-03	84544	AC007134 Arabidopsis thaliana
gb_pr2:ATAC007134	- 51.50	87.99	7.1e-03	203273	AC007134 Homo sapiens chrom
gb_pr2:ATAC007134	+ 51.00	137.28	12.67	243	U91332 Human cell line THAI DCH
gb_pr2:ATAC007134	+ 51.00	137.28	12.67	243	U91334 Human cell line THAI DCH
gb_pr2:ATAC007134	+ 51.00	137.28	12.67	243	U91336 Human cell line THAI DCH
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014769 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014771 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014773 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014775 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014777 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014779 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014781 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.19	12.82	246	AF014783 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.07	13.02	250	AF014785 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.07	13.02	250	AF014787 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	137.07	13.02	250	AF014789 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	136.95	13.22	234	U59965 Human MHC class I protei
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	D89333 Homo sapiens DNA for MHC
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ131852 Homo sapiens HLA-B gen
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ007603 Homo sapiens HLA-B*15
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ007605 Homo sapiens HLA-B*15
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	M28110 Human MHC class I cell
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AF002275 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AF016299 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67748 Human MHC class I antige
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U93913 Homo sapiens MHC class I
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ132713 Homo sapiens HLA-B gen
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ237594 Homo sapiens HLA-B gen
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ243433 Homo sapiens HLA-B gen
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ243737 Homo sapiens HLA-B gen
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AJ249317 Homo sapiens HLA-B gen
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67330 Human MHC class I antige
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AF017080 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67392 Human HLA-B gene, allele
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67390 Human HLA-B gene, allele
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67396 Human HLA-B gene, allele
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67396 Human HLA-B gene, allele
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AF002268 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	AF008926 Homo sapiens MHC class
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U67746 Human MHC class I antige
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U80670 Human class I MHC antige
gb_pr2:ATAC007134	+ 51.00	136.49	14.03	270	U90244 Human MHC class I antige

gb\_pr2:HSHLABV01 + 51.00 136.49 14.03 270 ! U55022 Human MHC class I ant  
 gb\_pr2:HSHLABV01 + 51.00 136.49 14.03 270 ! U91330 Homo sapiens MHC clas  
 gb\_pr4:HUM31HLA1 + 51.00 136.49 14.03 270 ! AF181842 Homo sapiens MHC cl

seq\_name: gb\_pl2:ATAC007134

## seq\_documentation\_block:

LOCUS ATAC007134 84544 bp DNA PLN 01-JUL-1999  
 DEFINITION Arabidopsis thaliana chromosome II BAC F7H1 genomic sequence,  
 complete sequence.

ACCESSION AC007134.8 GI:4678196

VERSION AC007134.8

KEYWORDS HTG.

SOURCE thale cress.

## ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.

1 (bases 1 to 84544)

Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,

Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.H., Benito,M.,

Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,

Fraser,C.M. and Venter,J.C.

Arabidopsis thaliana chromosome II BAC F7H1 genomic sequence

Unpublished

2 (bases 1 to 84544)

Lin,X. and Kaul,S.

Direct Submission

Submitted (21-MAR-1999) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org

3 (bases 1 to 84544)

Lin,X.

Direct Submission

Submitted (24-APR-1999) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

4 (bases 1 to 84544)

Lin,X.

Direct Submission

Submitted (18-JUN-1999) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

5 (bases 1 to 84544)

Lin,X.

Direct Submission

Submitted (01-JUL-1999) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

On Apr 24, 1999 this sequence version replaced gi:4580447.

Address all correspondence to:

Xiaoying Lin

The Institute for Genomic Research

9712 Medical Center Dr.

Rockville, MD 20850, USA

e-mail: xlin@tigr.org

BAC clone F7H1 is from Arabidopsis chromosome II and is near the

molecular marker m1398.

The orientation of the sequence is from SP6 to T7 end of the BAC

clone.

Genes were identified by a combination of three methods: Gene  
 prediction programs including GRAIL (available by anonymous ftp  
 from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of  
 Washington), Genscan (Chris Burge,  
 http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene  
 (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the  
 complete sequence against a peptide database and the Arabidopsis  
 EST database at TIGR (http://www.tigr.org/tdb/at.html).

Annotated genes are named to indicate the level of evidence for  
 their annotation. Genes with similarity to other proteins are named  
 after the database hits. Genes without significant peptide  
 similarity but with EST similarity are named as 'unknown' proteins.  
 Genes without protein or EST similarity, that are predicted by more  
 than two gene prediction programs over most of their length are  
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
 predicted by tRNAscan-SE (Sean Eddy).



```

78760 ATAAATAGATCAAGCAGAGCGTAAACCAATTGAAAAATATCTGGGTA 78809
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

20 x 20

78810 T 78810

seq\_name: gb\_htg6:AC010932

```

seq documentation_block:
LOCUS AC010932 203273 bp DNA HTG 30-NOV-1999
DEFINITION Homo sapiens chromosome 15 clone RP11-296E22 map 15, *** SEQUENCING
IN PROGRESS ***, 36 unordered pieces.
ACCESSION AC010932
VERSION AC010932.2 GI:6479157
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 203273)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
TITLE Homo sapiens chromosome 15, clone RP11-296E22
AUTHORS

```

```

REFERENCE 2 (bases 1 to 203273)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferrel,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tsaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

```

```

TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 30, 1999 this sequence version replaced gi:5931431.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L1684
Center clone name: 296_E22

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1375: contig of 1375 bp in length
* 1376 gap of unknown length
* 3332: contig of 1957 bp in length
* 3333 gap of unknown length
* 4414: contig of 1082 bp in length
* 4415 gap of unknown length
* 6085: contig of 1671 bp in length
* 6086 gap of unknown length
* 7636: contig of 1551 bp in length
* gap of unknown length

```

```

* 7637 9176: contig of 1540 bp in length
* gap of unknown length
* 9177 11451: contig of 2275 bp in length
* gap of unknown length
* 11452 14012: contig of 2561 bp in length
* gap of unknown length
* 14013 15805: contig of 1793 bp in length
* gap of unknown length
* 15806 17914: contig of 2109 bp in length
* gap of unknown length
* 17915 20644: contig of 2730 bp in length
* gap of unknown length
* 20645 23409: contig of 2765 bp in length
* gap of unknown length
* 23410 26283: contig of 2874 bp in length
* gap of unknown length
* 26284 28593: contig of 2310 bp in length
* gap of unknown length
* 28594 31564: contig of 2971 bp in length
* gap of unknown length
* 31565 34023: contig of 2459 bp in length
* gap of unknown length
* 34024 36890: contig of 2867 bp in length
* gap of unknown length
* 36891 41485: contig of 4595 bp in length
* gap of unknown length
* 41486 45667: contig of 4182 bp in length
* gap of unknown length
* 45668 49438: contig of 3771 bp in length
* gap of unknown length
* 49439 53273: contig of 3835 bp in length
* gap of unknown length
* 53274 56613: contig of 3340 bp in length
* gap of unknown length
* 56614 61902: contig of 5289 bp in length
* gap of unknown length
* 61903 67950: contig of 6048 bp in length
* gap of unknown length
* 67951 74487: contig of 6537 bp in length
* gap of unknown length
* 74488 82190: contig of 7703 bp in length
* gap of unknown length
* 82191 91448: contig of 9258 bp in length
* gap of unknown length
* 91449 98969: contig of 7521 bp in length
* gap of unknown length
* 98970 109518: contig of 10549 bp in length
* gap of unknown length
* 109519 119016: contig of 9498 bp in length
* gap of unknown length
* 119017 127923: contig of 8907 bp in length
* gap of unknown length
* 127924 139858: contig of 11935 bp in length
* gap of unknown length
* 139859 152524: contig of 12666 bp in length
* gap of unknown length
* 152525 164676: contig of 12152 bp in length
* gap of unknown length
* 164677 183255: contig of 18579 bp in length
* gap of unknown length
* 183256 203273: contig of 20018 bp in length.

```

```

FEATURES
Location/Qualifiers
1..203273
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-296E22"
/clone_lib="RPC1-11 Human Male BAC"

```

```

BASE COUNT 54651 a 48192 c 46411 g 53915 t 104 others
ORIGIN

```

```

alignment_scores:
  Quality: 51.50      Length: 23
  Ratio: 2.861        Gaps: 1
  Percent Similarity: 78.261  Percent Identity: 52.174

alignment_block:
  US-08-653-294-35 x AC010932/rev ..
  Align seg 1/1 to reverse of: AC010932 from: 1 to: 203273
  1 TyrglyArgLeuAsnArgLeuSerGluArgGlySer.....Le 14
  24858 TATGCCAGGAATGGGAGATTCTCAGATAAAGTACTCTCTGTACCT 24809
  14 uArgAsnLeuArgGlyTyr 20
  24808 CAAAATCTCAGAGGCTAC 24790

seq_name: gb_pr2:HSHLABK1

seq_documentation_block:
  LOCUS HSHLABK1 243 bp DNA PRI 26-MAR-1997
  DEFINITION Human cell line THAI DCH022 MHC class I HLA-B gene (allele
  HLA-B*1521), exon 2.
  ACCESSION U91332
  VERSION U91332.1 GI:1906671
  KEYWORDS
  SEGMENT
  SOURCE
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
  1 (bases 1 to 243)
  AUTHORS Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,
  Rungroung,E. and Bejjchandra,S.
  TITLE B15 alleles (B*1521)
  JOURNAL Unpublished
  AUTHORS
  TITLE Direct Submission
  JOURNAL Submitted (26-FEB-1997) Transfusion Medicine, Faculty of Medicine,
  Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
  Thailand
  FEATURES
    source
    1..243
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /map="6p21"
    /cell_type="lymphoblastoid"
    /cell_line="THAI DCH022"
    <1..243
    /gene="HLA-B"
    /number=2
  BASE COUNT 53 a 79 c 83 g 28 t
  ORIGIN
  alignment_scores:
    Quality: 51.00      Length: 10
    Ratio: 5.100        Gaps: 0
    Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-653-294-35 x HSHLABK1 ..
  Align seg 1/1 to: HSHLABK1 from: 1 to: 243
  11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20
  195 CGAGAGAGCCTCGGAACCTCGCGGCTAC 224

```

```

seq_name: gb_pr2:HSHLABM1

seq_documentation_block:
  LOCUS HSHLABM1 243 bp DNA PRI 26-MAR-1997
  DEFINITION Human cell line THAI DCH023 MHC class I HLA-B gene (allele
  HLA-B*1525), exon 2.
  ACCESSION U91334
  VERSION U91334.1 GI:1906675
  KEYWORDS
  SEGMENT
  SOURCE
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
  1 (bases 1 to 243)
  AUTHORS Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,
  Rungroung,E., Bejjchandra,S., Juji,I., Tokunaga,K. and
  Grosse-Wilde,H.
  TITLE B15 alleles (B*1525)
  JOURNAL Unpublished
  AUTHORS
  TITLE Direct Submission
  JOURNAL Submitted (26-FEB-1997) Transfusion Medicine, Faculty of Medicine,
  Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
  Thailand
  FEATURES
    Location/Qualifiers
    1..243
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /map="6p21"
    /cell_type="lymphoblastoid"
    /cell_line="THAI DCH023"
    1..243
    /gene="HLA-B"
    /notes="similar to exon 2 of B*1501"
    /number=2
  BASE COUNT 52 a 79 c 84 g 28 t
  ORIGIN
  alignment_scores:
    Quality: 51.00      Length: 10
    Ratio: 5.100        Gaps: 0
    Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-653-294-35 x HSHLABM1 ..
  Align seg 1/1 to: HSHLABM1 from: 1 to: 243
  11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20
  195 CGAGAGAGCCTCGGAACCTCGCGGCTAC 224

seq_name: gb_pr2:HSHLABO1

seq_documentation_block:
  LOCUS HSHLABO1 243 bp DNA PRI 26-MAR-1997
  DEFINITION Human cell line THAI DCH012 MHC class I HLA-B gene (allele
  HLA-B*1525), exon 2.
  ACCESSION U91336
  VERSION U91336.1 GI:1906679
  KEYWORDS
  SEGMENT
  SOURCE
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
  1 (bases 1 to 243)

```



**AUTHORS** Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,  
Rungroung,E., Bejchandra,S., Juji,T., Tokunaga,K. and  
Grosse-Wilde,H.  
**JOURNAL** B15 alleles (B\*1525)  
Unpublished  
**REFERENCE** 2 (bases 1 to 243)  
**AUTHORS** Chandanayingyong,D., Sirikong,M., Srinak,D., Longta,K.,  
Rungroung,E., Bejchandra,S., Juji,T., Tokunaga,K. and  
Grosse-Wilde,H.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (26-FEB-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand

**FEATURES**

## source

1. .243 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21"  
/cell\_line="THAI DCH012"  
/cell\_type="lymphoblastoid"

## exon

1. .243  
/gene="HLA-B"  
/note="Allele: HLA-B\*1525; similar to exon 2 of B\*1501"  
/number=2

**BASE COUNT** 52 a 79 c 84 g 28 t  
**ORIGIN**

**alignment\_scores:**

Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

**alignment\_block:**

US-08-653-294-35 x HSHLAB01 ..

Align seg 1/1 to: HSHLAB01 from: 1 to: 243

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
195 CGAGAGAGCTCGGAACCTGCGGCTAC 224

seq\_name: gb\_pr2:AF014769

seq\_documentation\_block: 246 bp DNA PRI 27-AUG-1997  
LOCUS AF014769  
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1502 allele),  
exon 2 and partial cds.

ACCESSION AF014769

VERSION AF014769.1 GI:2345103

**KEYWORDS**

SOURCE human.

**ORGANISM**

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**

1 (bases 1 to 246)  
**AUTHORS** Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.

**TITLE**

B15 alleles (B\*1502)

**JOURNAL**

Unpublished

**AUTHORS**

Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
Direct Submission

**TITLE**

Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand

**FEATURES**

## source

1. .246 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH4060"

/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
<1. .>246  
/gene="HLA-B"  
/allele="HLA-B\*1502"  
1. .246  
/gene="HLA-B"  
/note="2"  
/number=2  
<1. .>246  
/gene="HLA-B"  
/codon\_start=3  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67807.1"  
/db\_xref="GI:2345104"  
/translation="TAMSRGRGEPRIAYGVDDTQFVRFDSDAASPRMAPRAPWIE  
QEGPEYWDRTQISKNTQTYRSLRLRGYNQSEA"  
**BASE COUNT** 55 a 81 c 82 g 28 t  
**ORIGIN**

**alignment\_scores:**

Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

**alignment\_block:**

US-08-653-294-35 x AF014769 ..

Align seg 1/1 to: AF014769 from: 1 to: 246

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
198 CGAGAGAGCTCGGAACCTGCGGCTAC 227

seq\_name: gb\_pr2:AF014771

**seq\_documentation\_block:**

LOCUS AF014771 246 bp DNA PRI 27-AUG-1997  
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1502 allele),  
exon 2 and partial cds.

ACCESSION AF014771

VERSION AF014771.1 GI:2345107

**KEYWORDS**

SOURCE human.

**ORGANISM**

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**

1 (bases 1 to 246)

**AUTHORS**

Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.

**TITLE**

B15 alleles (B\*1502)

**JOURNAL**

Unpublished

**AUTHORS**

Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.

**TITLE**

Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand

**FEATURES**

## source

1. .246 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH4061"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"

## gene

<1. .>246  
/gene="HLA-B"

## exon

1. .246  
/allele="HLA-B\*1502"

```
/gene="HLA-B"
/note="2"
/number=2
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"
/protein_id="AB67809.1"
/db_xref="GI:2345108"
/translation="TAMSRPGRGEPRFIAVGVDVDTQFVRFDSDAASPRMAPRPWIE
QEGPEYWDNTQISKNTQTYRESLRLRGYINQSEA"
BASE COUNT      55 a      81 c      82 g      28 t
ORIGIN

alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-653-294-35 x AF014771 ..
Align seg 1/1 to: AF014771 from: 1 to: 246
11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20
|||||
198 CGAGAGAGCCTGCGGAACCTGCGGGCTAC 227

seq_name: gb_pr2:AF014771 ..
seq_documentation_block:
LOCUS AF014773      246 bp      DNA      PRI      27-AUG-1997
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*1502 allele),
exon 2 and partial cds.
ACCESSION AF014773
VERSION AF014773.1 GI:2345111
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 246)
AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
B15 alleles (B*1502)
Unpublished
JOURNAL
2 (bases 1 to 246)
REFERENCE Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
AUTHORS Direct Submission
TITLE Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
FEATURES
source
Location/Qualifiers
1..246
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Thai DCH3086"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
<1..>246
/gene="HLA-B"
/allele="HLA-B*1502"
1..246
/gene="HLA-B"
/number=2
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"
/protein_id="AB67813.1"
/db_xref="GI:2345116"
/translation="TAMSRPGRGEPRFIAVGVDVDTQFVRFDSDAASPRMAPRPWIE
QEGPEYWDNTQISKNTQTYRESLRLRGYINQSEA"
BASE COUNT      55 a      81 c      82 g      28 t
ORIGIN

/gene="HLA-B"
/note="2"
/number=2
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"
/protein_id="AB67809.1"
/db_xref="GI:2345108"
/translation="TAMSRPGRGEPRFIAVGVDVDTQFVRFDSDAASPRMAPRPWIE
QEGPEYWDNTQISKNTQTYRESLRLRGYINQSEA"
BASE COUNT      55 a      81 c      82 g      28 t
ORIGIN

alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-653-294-35 x AF014771 ..
Align seg 1/1 to: AF014771 from: 1 to: 246
11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20
|||||
198 CGAGAGAGCCTGCGGAACCTGCGGGCTAC 227

seq_name: gb_pr2:AF014773 ..
seq_documentation_block:
LOCUS AF014773      246 bp      DNA      PRI      27-AUG-1997
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*1502 allele),
exon 2 and partial cds.
ACCESSION AF014773
VERSION AF014773.1 GI:2345111
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 246)
AUTHORS Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
B15 alleles (B*1502)
Unpublished
JOURNAL
2 (bases 1 to 246)
REFERENCE Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,
Rungroung,E. and Bejchandra,S.
AUTHORS Direct Submission
TITLE Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
Thailand
FEATURES
source
Location/Qualifiers
1..246
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Thai DCH3086"
/chromosome="6"
/map="6p21"
/cell_type="lymphoblastoid"
<1..>246
/gene="HLA-B"
/allele="HLA-B*1502"
1..246
/gene="HLA-B"
/number=2
<1..>246
/gene="HLA-B"
/codon_start=3
/product="MHC class I antigen HLA-B"
/protein_id="AB67813.1"
/db_xref="GI:2345116"
/translation="TAMSRPGRGEPRFIAVGVDVDTQFVRFDSDAASPRMAPRPWIE
QEGPEYWDNTQISKNTQTYRESLRLRGYINQSEA"
BASE COUNT      55 a      81 c      82 g      28 t
ORIGIN
```

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-35 x AF014775 ..

Align seg 1/1 to: AF014775 from: 1 to: 246

11 ArgGlusSerLeuArgAsnLeuArgGlyTyr 20  
 |||  
 198 CGAGAGAGCCTGCGGAACCTGCGGCTAC 227

seq\_name: gb\_pr2:AF014777

seq\_documentation\_block: 246 bp DNA PRI 27-AUG-1997  
 LOCUS AF014777  
 DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1502 allele),  
 exon 2 and partial cds.  
 ACCESSION AF014777  
 VERSION AF014777.1 GI:2345119  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 246)  
 AUTHORS Chananayongyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungroung,E. and Bejchandra,S.  
 TITLE B15 alleles (B\*1502)  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 246)  
 AUTHORS Chananayongyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungroung,E. and Bejchandra,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
 Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
 Thailand

FEATURES Location/Qualifiers

source  
 1..246  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="Thai DCH017"  
 /chromosome="6"  
 /map="6p21"  
 /cell\_type="lymphoblastoid"  
 <1..>246  
 /gene="HLA-B"  
 /allele="HLA-B\*1502"  
 1..246  
 /gene="HLA-B"  
 /note="2"  
 <1..>246  
 /gene="HLA-B"  
 /codon\_start=3  
 /product="MHC class I antigen HLA-B"  
 /protein\_id="AAB67815.1"  
 /db\_xref="GI:2345120"  
 /translation="TAMSRPGEGPRFIAVGVDQTQVRFDSDAASPRMAPRAPWIE  
 QEGPEYDNRNTQISKNTQTYRESLRLRGYNQSEA"  
 55 a 81 c 82 g 28 t

BASE COUNT  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-35 x AF014777 ..

Align seg 1/1 to: AF014777 from: 1 to: 246

11 ArgGlusSerLeuArgAsnLeuArgGlyTyr 20  
 |||  
 198 CGAGAGAGCCTGCGGAACCTGCGGCTAC 227

seq\_name: gb\_pr2:AF014779

seq\_documentation\_block: 246 bp DNA PRI 27-AUG-1997  
 LOCUS AF014779  
 DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1502 allele),  
 exon 2 and partial cds.  
 ACCESSION AF014779  
 VERSION AF014779.1 GI:2345123  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 246)  
 AUTHORS Chananayongyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungroung,E. and Bejchandra,S.  
 TITLE B15 alleles (B\*1502)  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 246)  
 AUTHORS Chananayongyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
 Rungroung,E. and Bejchandra,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
 Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
 Thailand

FEATURES Location/Qualifiers

source  
 1..246  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="Thai DCH002"  
 /chromosome="6"  
 /map="6p21"  
 /cell\_type="lymphoblastoid"  
 <1..>246  
 /gene="HLA-B"  
 /allele="HLA-B\*1502"  
 1..246  
 /gene="HLA-B"  
 /note="2"  
 <1..>246  
 /gene="HLA-B"  
 /codon\_start=3  
 /product="MHC class I antigen HLA-B"  
 /protein\_id="AAB67817.1"  
 /db\_xref="GI:2345124"  
 /translation="TAMSRPGEGPRFIAVGVDQTQVRFDSDAASPRMAPRAPWIE  
 QEGPEYDNRNTQISKNTQTYRESLRLRGYNQSEA"  
 55 a 81 c 82 g 28 t

BASE COUNT  
 ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-35 x AF014779 ..

Align seg 1/1 to: AF014779 from: 1 to: 246

11 ArgGlusSerLeuArgAsnLeuArgGlyTyr 20  
 |||  
 198 CGAGAGAGCCTGCGGAACCTGCGGCTAC 227

[illegible]

REFERENCE  
AUTHORS 1 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE B15 alleles (B\*1525)  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand  
FEATURES  
source Location/Qualifiers  
1..250  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH1109"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
1..250  
/gene="HLA-B"  
/allele="HLA-B\*1525"  
/note="2"  
/number=2  
/gene="HLA-B"  
/codon\_start=1  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67823.1"  
/db\_xref="GI:2345136"  
/translation="FYTAMSRPGRGEPRFIAVGYYDDTQFVRFDSDAASPRMAPRAPW  
IEQEGPEYWDRETOISKNTQTYRESLNLRGYYNQSEA"  
BASE COUNT 54 a 81 c 84 g 31 t  
ORIGIN  
alignment\_scores  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-653-294-35 x AF014785 ..  
Align seg 1/1 to: AF014785 from: 1 to: 250  
11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
202 CGAGAGAGCCTCGGACCTGCGCGGTAC 231  
seq\_name: gb\_pr2:AF014787  
seq\_documentation\_block:  
LOCUS AF014787 250 bp DNA PRI 27-AUG-1997  
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1525 allele),  
exon 2 and partial cds.  
ACCESSION AF014787  
VERSION AF014787.1 GI:2345139  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE B15 alleles (B\*1525)  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand  
FEATURES  
source Location/Qualifiers  
1..250  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH1109"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
1..250  
/gene="HLA-B"  
/allele="HLA-B\*1525"  
/note="2"  
/number=2  
/gene="HLA-B"  
/codon\_start=1  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67823.1"  
/db\_xref="GI:2345136"  
/translation="FYTAMSRPGRGEPRFIAVGYYDDTQFVRFDSDAASPRMAPRAPW  
IEQEGPEYWDRETOISKNTQTYRESLNLRGYYNQSEA"  
BASE COUNT 54 a 81 c 84 g 31 t  
ORIGIN  
alignment\_scores  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-653-294-35 x AF014785 ..  
Align seg 1/1 to: AF014785 from: 1 to: 250  
11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
202 CGAGAGAGCCTCGGACCTGCGCGGTAC 231  
seq\_name: gb\_pr2:AF014787  
seq\_documentation\_block:  
LOCUS AF014787 250 bp DNA PRI 27-AUG-1997  
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B\*1525 allele),  
exon 2 and partial cds.  
ACCESSION AF014787  
VERSION AF014787.1 GI:2345139  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE B15 alleles (B\*1525)  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand  
FEATURES  
source Location/Qualifiers  
1..250  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH1109"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
1..250  
/gene="HLA-B"  
/allele="HLA-B\*1525"  
/note="2"  
/number=2  
/gene="HLA-B"  
/codon\_start=1  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67825.1"  
/db\_xref="GI:2345140"  
/translation="FYTAMSRPGRGEPRFIAVGYYDDTQFVRFDSDAASPRMAPRAPW  
IEQEGPEYWDRETOISKNTQTYRESLNLRGYYNQSEA"  
BASE COUNT 54 a 81 c 84 g 31 t  
ORIGIN  
alignment\_scores  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-653-294-35 x AF014787 ..  
Align seg 1/1 to: AF014787 from: 1 to: 250  
11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
202 CGAGAGAGCCTCGGACCTGCGCGGTAC 231

REFERENCE  
AUTHORS 1 (bases 1 to 250)  
Chandanayingyong,D., Sirikong,M., Luangtrakool,K., Srinak,D.,  
Rungroung,E. and Bejchandra,S.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1997) Transfusion Medicine, Faculty of Medicine,  
Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,  
Thailand  
FEATURES  
source Location/Qualifiers  
1..250  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Thai DCH3258"  
/chromosome="6"  
/map="6p21"  
/cell\_type="lymphoblastoid"  
1..250  
/gene="HLA-B"  
/allele="HLA-B\*1525"  
/note="2"  
/number=2  
/gene="HLA-B"  
/codon\_start=1  
/product="MHC class I antigen HLA-B"  
/protein\_id="AAB67825.1"  
/db\_xref="GI:2345140"  
/translation="FYTAMSRPGRGEPRFIAVGYYDDTQFVRFDSDAASPRMAPRAPW  
IEQEGPEYWDRETOISKNTQTYRESLNLRGYYNQSEA"  
BASE COUNT 54 a 81 c 84 g 31 t  
ORIGIN  
alignment\_scores  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-653-294-35 x AF014787 ..  
Align seg 1/1 to: AF014787 from: 1 to: 250  
11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
202 CGAGAGAGCCTCGGACCTGCGCGGTAC 231

**THIS PAGE BLANK (USPTO)**



FT intron complement (1412..1560)  
 FT /\*tag= d  
 FT /note= "SV40 small t intron"  
 FT 3'utr complement (1561..1794)  
 FT /\*tag= e  
 FT /note= "3' untranslated region of HLA-B7  
 FT heavy chain mRNA"  
 FT cds complement (1795..2880)  
 FT /\*tag= f  
 FT /note= "HLA-B7 open reading frame"  
 FT ltr complement (2886..3415)  
 FT /\*tag= g  
 FT /note= "Rous sarcoma virus 3' LTR promoter region"  
 FT misc\_feature 3416..4059  
 FT /\*tag= h  
 FT /note= "pBR322 backbone"  
 PN WO9429469-A.  
 PD 22-DEC-1994.  
 PE 27-MAY-1994; U06069.  
 PR 07-JUN-1993; US-074344.  
 PA (UNMI ) UNIV MICHIGAN.  
 PA (VICA-) VICAL INC.  
 PI Lew D, Marquet M, Nabel EG, Nabel GJ;  
 DR WPI; 95-036494/05.  
 PT New vectors for gene therapy, partic for tumours - comprising  
 PT genetic material encoding one or more cistron(s) which express  
 PT immunogenic or therapeutic peptide(s)  
 PS Claim 9; Page 42-43; 50pp; English.  
 CC This HLA-B7 antigen encoding plasmid was developed to incorporate many  
 CC advantageous features, eg. the kanamycin resistance gene. The  
 CC eradication of two open reading frames encoding portions of SV40 viral  
 CC proteins lowers the risk of tumorigenicity. The vector may also operate  
 CC as a cassette into which cistrons may be inserted and removed at will  
 CC for the transcription and subsequent translation of peptides of interest.  
 CC The vector is used partic. for the treatment of neoplastic disease,  
 CC eg. melanoma, and provides enhanced gene delivery and expression  
 CC in vivo.  
 SQ Sequence 4059 BP; 975 A; 1051 C; 1033 G; 1000 T;

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-653-294-35 x Q75974/rev ..  
 Align seg 1/1 to reverse of: Q75974 from: 1 to: 4059

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
 |||||  
 2589 CGAGAGAGCGTCGCGACCTGCGCGGCTAC 2560

seq\_name: N\_Geneseq\_36:Q75973

seq\_documentation\_block:  
 ID Q75973 standard; CDNA; 4965 BP.  
 AC Q75973;  
 DE 23-AUG-1995 (first entry)  
 DE pHLA-B7/beta-2 microglobulin expression vector.  
 KW expression vector; pHLA-B7/beta-2 microglobulin; heavy human HLA-B7;  
 KW light beta-2 microglobulin; class I major histocompatibility complex;  
 KW MHC; bicistronic mRNA; human leukocyte antigen; HLA;  
 KW covalently closed circular DNA; ds.  
 OS Synthetic.  
 FH Key  
 FH Location/Qualifiers  
 FT 1..529  
 FT ltr  
 FT /\*tag= a  
 FT /note= "Rous sarcoma virus LTR promoter domain, derived  
 FT for the Schmidt-Rupin strain nucleotides  
 FT 8673-9146. This region also includes a 56 bp  
 FT region of a synthetic oligonucleotide which

FT modifies this regulatory sequence to effect a  
 FT higher level of expression of downstream  
 FT sequences. The oligonucleotide removes a  
 FT polyadenylation signal sequence originally found  
 FT in the RSV DNA sequence."  
 FT 531..534  
 FT /\*tag= b  
 FT /label= consensus\_Kozak\_signal\_sequence  
 FT 535..1620  
 FT /\*tag= c  
 FT /note= "HLA-B7 heavy chain open reading frame"  
 FT 535..606  
 FT /\*tag= d  
 FT /note= "encodes putative signal peptide of the  
 FT HLA-B7 heavy chain"  
 FT 607..1620  
 FT /\*tag= e  
 FT /note= "encodes putative HLA-B7 heavy chain mature  
 FT peptide"  
 FT 1621..1853  
 FT /\*tag= f  
 FT /note= "3' untranslated sequence of HLA-B7 heavy  
 FT chain mRNA"  
 FT 1854..1888  
 FT /\*tag= g  
 FT /note= "multiple cloning site, forms a junction  
 FT between the HLA-B7 sequence and the EMCV-CITE  
 FT sequence, and is used to facilitate subcloning"  
 FT 1889..2479  
 FT /\*tag= h  
 FT /note= "murine encephalomyocarditis CAP-independent  
 FT translational enhancer (EMCV-CITE); taken from  
 FT nucleotides 255-843 of cloned EMCV genomic DNA.  
 FT It is a non-coding regulatory sequence functioning  
 FT as an internal entry point for the eukaryotic  
 FT ribosomal subunits when located within a mRNA  
 FT mol. . It enables the translational start codon of  
 FT the beta-2 microglobulin, downstream of the HLA-B7  
 FT stop codon on this bicistronic mRNA to be  
 FT recognised by the ribosome"  
 FT 2480..2839  
 FT /\*tag= i  
 FT /note= "encodes beta-2 microglobulin; this cDNA is  
 FT deriv. from chimpanzee (differs to the human  
 FT cDNA by only 4 bases)"  
 FT 2840..2846  
 FT /\*tag= j  
 FT /note= "3' untranslated region of the beta-2  
 FT microglobulin mRNA"  
 FT 2847..2870  
 FT /\*tag= k  
 FT /note= "synthetic linker"  
 FT 3'UTR ..3111  
 FT /\*tag= l  
 FT /note= "bovine growth hormone 3'UTR and transcriptional  
 FT terminator; it starts at a blunt-ended BglI site  
 FT within the 3'UTR of the mRNA coding sequence"  
 FT 2979..2984  
 FT /\*tag= m  
 FT 3112..3151  
 FT /\*tag= n  
 FT /note= "synthetic linker to facilitate cloning"  
 FT complement (3151..3967)  
 FT /\*tag= o  
 FT /note= "kanamycin resistance gene open reading frame;  
 FT the gene is taken from the transposable element  
 FT Tn903"  
 FT 4014..4965  
 FT /\*tag= p  
 FT /note= "pBR322 backbone contg. bacterial origin of  
 FT replication, it represents nucleotides 2244-3193"  
 PN WO9429469-A.  
 PD 22-DEC-1994.



PF 27-MAY-1994; U06069.  
 PR 07-JUN-1993; US-074344.  
 PA (UNMI ) UNIV MICHIGAN.  
 PA (VICA-) VICAL INC.  
 PI Lew D, Marquet M, Nabel EG, Nabel GU;  
 DR WPI: 95-036494/05.  
 PT New vectors for gene therapy, partic for tumours - comprising  
 PT genetic material encoding one or more cistron(s) which express  
 PT immunogenic or therapeutic peptide(s)  
 PS Claim 8; Page 41-42; 50pp; English.  
 CC The pHLA-B7/beta-2 microglobulin plasmid expression vector, in addition  
 CC to the kanamycin resistance gene, contains the plasmid DNA encoding the  
 CC heavy (human HLA-B7) and light (beta-2 microglobulin) proteins of a  
 CC class I major histocompatibility complex (MHC) antigen. The plasmid is  
 CC designed to express these two proteins via a bicistronic mRNA in  
 CC eukaryotic cells. Initiation of transcription of the mRNA is dependent  
 CC on a Rous Sarcoma Virus (RSV) promoter sequence deriv. from the 3' long  
 CC terminal repeat. Termination of transcription is dependent upon the  
 CC polyadenylation signal sequence deriv. from the bovine growth hormone  
 CC gene. Eukaryotic cell translation of the heavy chain is regulated by the  
 CC 5' cap-dependent protein start site. Translation of the light chain is  
 CC controlled by the CITE. Finally the replication of the plasmid in  
 CC bacterial cells is controlled by the presence of a bacterial origin of  
 CC replication. The vector is used partic. for the treatment of neoplastic  
 CC disease, eg. melanoma, and provides enhanced gene delivery and expression  
 CC in vivo.  
 SQ Sequence 4965 BP; 1171 A; 1293 C; 1338 G; 1163 T;

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-653-294-35 x Q75973 ..

Align seg 1/1 to: Q75973 from: 1 to: 4965

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
 |||||  
 829 CGAGAGACCTGGCGAACCTGGCGGGCTAC 858

seq\_name: N\_Geneseq\_36:Q22494

seq\_documentation\_block:  
 ID Q22494 standard; DNA; 2675 BP.  
 DT 28-AUG-1992 (first entry)  
 DE Human CRIPTO-related gene CR-3.  
 KW cell proliferation; tumour; CR-1; transforming growth factor;  
 KW epidermal growth factor; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 809..1375  
 FT /\*tag= a  
 FT /product= CR-3  
 PN US7749001-A.  
 PD 25-FEB-1992.  
 PF 23-AUG-1991; 749001.  
 PR 23-AUG-1991; US-749001.  
 PA (USSH ) US DEPT HEALTH & HUMAN.  
 PI Salomon D, Persico M;  
 DR WPI: 92-123675/15.  
 DR P-PSDB; R22347.  
 PT New CRIPTO gene CR-1 and CRIPTO-related gene CR-3 genomic DNA -  
 PT CR-3 protein and anti-CR-3 antibodies, useful in immunoassay to  
 PT detect CR-3 as tumour specific marker  
 PS Disclosure; Page 29; 44pp; English.  
 CC The CR-3 nucleotide sequence was isolated from a human genomic  
 CC library using a labelled DNA fragment containing 800bp upstream of  
 CC the translation initiation site of CR-1. Clones were isolated which  
 CC contained the 5' cDNA non-coding region of CR-3. One clone was

CC sequenced and found to include a complete CRIPTO cDNA lacking  
 CC introns and containing a poly(A) tract at the 3' end. Seven single  
 CC base pair substitutions were observed in the coding region (c.f.  
 CC CR-1 sequence in Ciccodicola A. et al., EMBO J.8:1987-1991(1989)).  
 CC six of which gave rise to amino acid changes. The 3' non-coding  
 CC sequence is less similar (97% identical) to CR-1. Most of the  
 CC differences occur in the inverted Alu sequence. The unusual poly(A)  
 CC addition site AGTAAA found in the CR-1 gene is also conserved in  
 CC CR-3. The similarity between CR-1 and CR-3 extends for 697  
 CC nucleotides upstream of the initiator AUG where it is possible to  
 CC observe 7 base pair substitutions and 6 nucleotide deletions.  
 CC The 14 "others" in the sequence represent nucleotides whose  
 CC identity, although determined by the inventors, was unclear in  
 CC the sequence printed in the specification.  
 SQ Sequence 2675 BP; 738 A; 611 C; 553 G; 759 T;

alignment\_scores:  
 Quality: 48.00 Length: 21  
 Ratio: 3.200 Gaps: 1  
 Percent Similarity: 71.429 Percent Identity: 57.143

alignment\_block:  
 US-08-653-294-35 x Q22494 ..

Align seg 1/1 to: Q22494 from: 1 to: 2675

2 GlyArgLeuAsnArgLeuSerGlu.....ArgArgGluSerLeuArgAs 16  
 |||||  
 2161 GGGAGACTGGGTAGGAAGAGCAATAAGAGAGAGAGAGAGAGAGAGAGAG 2210

16 nLeuArgGlyTyr 20  
 |||||  
 2211 CAAATGGGTAC 2223

seq\_name: N\_Geneseq\_36:T71145

seq\_documentation\_block:  
 ID T71145 standard; DNA; 2675 BP.  
 DT 19-AUG-1997 (first entry)  
 DE DNA encoding human CRIPTO-related gene product, CR-3.  
 KW CRIPTO; CR1; CR3; epidermal growth factor superfamily; EGF; mitogen;  
 KW tumour marker; epithelium; mesenchyme; diagnosis; prognosis; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 809..1375  
 FT /\*tag= a  
 FT /product= CR-3  
 PN US5620866-A.  
 PD 15-APR-1997.  
 PF 23-AUG-1991; 749001.  
 PR 23-AUG-1991; US-749001.  
 PR 17-NOV-1993; US-154198.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Persico MG, Salomon DS;  
 DR WPI: 97-235180/21.  
 DR P-PSDB; W19980.  
 PT New DNA encoding the human CRIPTO-related gene product CR-3 - used  
 PT to produce recombinant protein, useful for preparation of diagnostic  
 PT antibodies  
 PS Claim 1; Column 17-20; 25pp; English.  
 CC T71145 encodes a human CRIPTO-related gene product CR-3. CR-3 is  
 CC thought to be a tumour-specific marker and as such DNA and amino  
 CC acid sequences of CR-3 and derivatives can be used in the  
 CC diagnosis, prognosis and possibly treatment of some types of cancer.  
 CC CR-3 includes a region homologous to products of the epidermal  
 CC growth factor superfamily of genes and it may be a mitogen involved  
 CC in regulating proliferation, differentiation and transformation of  
 CC mesenchymal and epithelial cells.  
 SQ Sequence 2675 BP; 739 A; 618 C; 555 G; 763 T;





PI Persica MG, Salomon DS;  
 DR WPI: 99-095001/08.  
 DR P-PSDB: W87630.  
 PT New antibody to human CRIPTO-related polypeptide-3 - used for the  
 PT detection of CRIPTO-related polypeptide-3 which is involved in the  
 PT regulation of the proliferation, differentiation and transformation  
 PT of cells  
 PS Example 2: Columns 9-16: 26pp; English.  
 CC The present sequence encodes a human CRIPTO-related protein CR-1. The  
 CC specification describes an antibody which has binding affinity to CR-3  
 CC and not to CR-1. The antibodies can be used for the detection of CR-3.  
 CC CR-3 is a regulatory molecule involved in regulating the proliferation,  
 CC differentiation, and transformation of various mesenchymal and epithelial  
 CC cells. In addition expression of CR-3 may serve as a tumour specific  
 CC marker that may have applicability in the diagnosis, prognosis and  
 CC possible treatment of specific types of cancer.  
 SQ Sequence 5761 BP: 1529 A; 1304 C; 1229 G; 1699 T;

alignment\_scores:  
 Quality: 48.00 Length: 21  
 Ratio: 3.200 Gaps: 1  
 Percent Similarity: 71.429 Percent Identity: 57.143

alignment\_block:  
 US-08-653-294-35 x V83923 ..

Align seg 1/1 to: V83923 from: 1 to: 5761

2 GlyArgLeuAsnArgLeuSerGlu.....ArgArgGluSerLeuArgAs 16  
 |||||.....:|||||  
 4960 GGGAGACTGGGTAGGAAGAGGAGCAATAGAGAGAGAGGTTGAAAAA 5009

16 nLeuArgGlyTyr 20  
 | |||||  
 5010 CAAATGGGTTAC 5022

seq\_name: N\_Geneseq\_36:V17562

seq\_documentation\_block:  
 ID V17562 standard; CDNA; 1818 BP.  
 AC V17562;  
 DT 10-JUN-1998 (first entry)  
 DE Coding sequence for the alpha subunit of beta-conglycinin.  
 KW Beta-conglycinin; soybean seed protein; transgenic plant;  
 KW seed storage protein profile; ss.  
 OS Glycine max.  
 PN W09747731-A2.  
 PD 18-DEC-1997.  
 PF 10-JUN-1997; U09743.  
 PR 14-JUN-1996; US-019940.  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PI Fader GM, Kinney AJ;  
 DR WPI: 98-052298/05.  
 PT Suppression of specific classes of soybean seed protein genes -  
 PT useful to change seed storage protein profiles of transgenic plants  
 PS disclosure: Page 29-30: 58pp; English.  
 CC This sequence represents the coding sequence for the alpha subunit of the  
 CC soybean seed protein beta-conglycinin. The method of the invention is for  
 CC reducing the quantity of a soybean seed storage protein (A), such as  
 CC beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric  
 CC gene comprising: (i) a nucleic acid fragment encoding a promoter that is  
 CC functional in the cells of soybean seeds; (ii) a nucleic acid fragment  
 CC encoding all or a portion of (A) placed in sense or antisense orientation  
 CC relative to the promoter of (i); and (iii) a transcriptional termination  
 CC region; (b) creating a transgenic soybean cell by introducing into a  
 CC soybean cell the chimeric gene of (a); and (c) growing the transgenic  
 CC soybean cells of (b) under conditions that result in expression of the  
 CC chimeric gene of (a); where the quantity of one or more members of a  
 CC class of (A) subunits is reduced when compared to soybeans not containing  
 CC the chimeric gene of (a). The method is used to construct transgenic  
 CC soybean lines where the expression of genes encoding (A) are modulated to  
 CC effect a change in seed storage protein profile of transgenic plants.

CC Modification of the seed storage protein profile can result in the  
 CC production of novel soy protein products with unique and valuable  
 CC functional characteristics.  
 SQ Sequence 1818 BP: 581 A; 424 C; 427 G; 386 T;

alignment\_scores:  
 Quality: 47.50 Length: 23  
 Ratio: 2.794 Gaps: 1  
 Percent Similarity: 73.913 Percent Identity: 47.826

alignment\_block:  
 US-08-653-294-35 x V17562 ..

Align seg 1/1 to: V17562 from: 1 to: 1818

1 TyrGlyArg.....LeuAsnArgLeuSerGluArgGluSerLe 14  
 |||||.....:|||||  
 631 TATGTCGCATTTCGCTCCAGAGGTTCAACCCAGCTCCCAACT 680

14 uArgAsnLeuArgGlyTyr 20  
 |::::|||  
 681 TCAGATCTCCGAGACTAC 699

seq\_name: N\_Geneseq\_36:X12977

seq\_documentation\_block:  
 ID X12977 standard; DNA; 1346 BP.  
 AC X12977;  
 DT 19-MAR-1999 (first entry)  
 DE Enterococcus faecalis genome contig SEQ ID NO:40.  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.  
 OS Enterococcus faecalis.  
 PN W09850555-A2.  
 PD 12-NOV-1998.  
 PF 04-MAY-1998; U08985.  
 PR 14-NOV-1997; US-066009.  
 PR 06-MAY-1997; US-044031.  
 PR 16-MAY-1997; US-046655.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Dillon PJ, Kunsch CA;  
 DR WPI: 99-045171/04.  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 PS Claim 1; Page 412-413; 2084pp; English.  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC X12938 to X13919 represent these nucleotide sequences which are primary  
 CC nucleotide sequences, also known as contigs. The computer-based system  
 CC can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 SQ Sequence 1346 BP: 395 A; 269 C; 284 G; 395 T;

alignment\_scores:  
 Quality: 47.00 Length: 19  
 Ratio: 2.938 Gaps: 0  
 Percent Similarity: 84.211 Percent Identity: 42.105

alignment\_block:  
 US-08-653-294-35 x X12977/rev ..

Align seg 1/1 to reverse of: X12977 from: 1 to: 1346

```

1  TyrGlyArgLeuAsnArgLeuSerGluArgGluSerLeuArgAsnLe 17
|||||
1202 TATGGAGATTAGAGAAAGTCTGCATTAAAGACAGATTCGAAATTT 1153
17 uArgGly 19
   ::::|
1152 TAAAGGC 1146

```

seq\_name: N\_Geneseq\_36:V89291

seq\_documentation\_block:

```

ID V89291 standard; cDNA; 289 BP.
AC V89291;
DT 15-FEB-1999 (first entry)
DE EST clone CG336.
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN W09843436-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06955.
PR 10-APR-1997; US-838821.
PA (GEMV ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 99-070077/06.
SQ New polynucleotides encoding human secreted proteins - derived from
e.g. human blood, kidney, foetal lung, placenta, testes, brain,
ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1: Page 171: 618pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
are predicted to have useful biological activities which would make
them suitable for treating, preventing or ameliorating medical
conditions in humans and animals, although no supporting data is
given. Suggested activities include nutritional activity, immune
stimulating or suppressing activity, haematopoiesis regulating
activity, tissue growth activity, activin/inhibin activity,
chemotactic/chemokinetic activity, haemostatic and thrombolytic
activity, receptor/ligand activity, anti-inflammatory activity,
cadherin/tumour invasion suppressor activity, tumour inhibition
activity. The polynucleotide may also be useful for gene therapy.
SQ Sequence 289 BP; 63 A; 63 C; 63 G; 100 T;

```

alignment\_scores:

```

Quality: 46.00 Length: 13
Ratio: 4.182 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 76.923

```

alignment\_block:

US-08-653-294-35 x V89291 ..

Align seg 1/1 to: V89291 from: 1 to: 289

```

3 ArgLeuAsnArgLeuSerGluArgGluSerLeuArg 15
|||||
74 AGAGTCACAGCGTAAAGTGCAGAGAGAGGAGGATGCTACGC 112

```

seq\_name: N\_Geneseq\_36:V82783

seq\_documentation\_block:

```

ID V82783 standard; cDNA; 745 BP.
AC V82783;
DT 25-FEB-1999 (first entry)
DE Clone dn809.5 isolated from human foetal brain cDNA library.
KW Secreted protein; nutritional activity; immune stimulating; vaccine;
KW suppressing activity; haematopoiesis regulating activity;
KW tissue growth activity; activin; inhibin activity; chemotactaxis;

```

```

KW chemokinetic activity; haemostasis; thrombolytic activity; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy; ds.
OS Homo sapiens.
PN W09842739-A2.
PD 01-OCT-1998.
PF 20-MAR-1998; U05653.
PR 19-MAR-1998; US-044466.
PR 21-MAR-1997; US-822167.
PA (GEMV ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-609890/51.
DR P-PSDB: W85461.
PT New polynucleotides encoding secreted human proteins - derived from
human foetal brain, adult brain, foetal kidney, placenta or adult
pituitary gland cDNA libraries.
PS Claim 26; Page 80; 119pp; English.
CC The present sequence encodes a secreted protein. The polynucleotide and
secreted protein are predicted to have biological activities which would
make them suitable for treating, preventing or ameliorating medical
conditions in humans and animals, although no supporting data is given.
Suggested activities include nutritional activity, immune stimulating
(e.g. as vaccines) or suppressing activity, haematopoiesis regulating
activity, tissue growth activity, activin/inhibin activity,
chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
invasion suppressor activity, and tumour inhibition activity (no data is
given in the specification to support these activities). The
polynucleotide is also stated to be useful for gene therapy.
SQ Sequence 745 BP; 179 A; 229 C; 176 G; 161 T;

```

alignment\_scores:

```

Quality: 46.00 Length: 13
Ratio: 4.182 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 76.923

```

alignment\_block:

US-08-653-294-35 x V82783/rev ..

Align seg 1/1 to reverse of: V82783 from: 1 to: 745

```

3 ArgLeuAsnArgLeuSerGluArgGluSerLeuArg 15
|||||
123 AGAGTCACAGCGTAAAGTGCAGAGAGAGGAGGATGCTACGC 85

```

seq\_name: N\_Geneseq\_36:T99594

seq\_documentation\_block:

```

ID T99594 standard; cDNA; 1849 BP.
AC T99594;
DT 22-JUN-1998 (first entry)
DE Mouse TIE-2 receptor ligand 3 cDNA.
KW TIE-2 receptor ligand; TIE ligand-3; tyrosine kinase; mouse;
KW angiogenesis; vascularisation; blood vessel growth; tumour;
KW therapy; receptorbody; ss.
OS Mus musculus.
PH Key Location/Qualifiers
FT CDS 47..1576
FT /tag= a
PN W09748804-A2.
PD 24-DEC-1997.
PF 19-JUN-1997; U10728.
PR 02-AUG-1996; US-022999.
PR 19-JUN-1996; US-665926.
PR 02-JUL-1996; US-021087.
PA (REGE-) REGENERON PHARM INC.
PI Jones PF, Valenzuela DM, Yancopoulos GD;
DR WPI: 98-063143/06.
DR P-PSDB: W26792.
PT DNA encoding TIE ligands 3 and 4 - useful for, e.g. blocking blood
vessel growth and promoting neovascularisation, etc.

```

PS Claim 23: Fig 6A-B; 86pp; English.  
 CC This cDNA sequence includes a coding region for novel mouse TIE  
 CC ligand-3 (mFL-4, see W26792), a TIE-2 receptor ligand that is a  
 CC tyrosine kinase with immunoglobulin and epidermal growth factor  
 CC homology domains. It was obtained by PCR amplification of cDNA  
 CC derived from mouse cell lines C2C12ras and M87 using primers  
 CC (see T99595-96) specific for mFL-3. A host-vector system for the  
 CC production of mFL-3 is claimed, as well as an antibody which  
 CC specifically binds mFL-3, a receptorbody which specifically binds  
 CC mFL-3, a conjugate comprising mFL-3 and a cytotoxic agent, and an  
 CC isolated nucleic acid (see T99593) coding for human TIE ligand 4  
 CC (see W26791). TL3, TL4, their antagonists, etc. are useful for  
 CC blocking blood vessel growth, promoting neovascularisation,  
 CC promoting or blocking the growth or differentiation of a cell  
 CC expressing the TIE receptor and attenuating or preventing tumour  
 CC growth.  
 SQ Sequence 1849 BP; 451 A; 549 C; 507 G; 342 T;

alignment\_scores:  
 Quality: 46.00 Length: 18  
 Ratio: 2.875 Gaps: 0  
 Percent Similarity: 88.889 Percent Identity: 50.000

alignment\_block:  
 US-08-653-294-35 x T99594 ..

Align seg 1/1 to: T99594 from: 1 to: 1849

3 ArgLeuAsnArgLeuSerGluArgGluSerLeuArgGlnLeuArgG1 19  
 ::::||||| :||:::|||||:|||||:|||||:||||| :||  
 707 CAGCTTACAGCCTCCACGAGAGGAGGCAACACTGCACAGTCTCCTGGG 756

19 yTyr 20

||:::

757 CCAT 760

seq\_name: N\_Geneseq\_36:T11549

seq\_documentation\_block:

ID T11549 standard; DNA; 13585 BP.

AC T11549;

DT 18-APR-1996 (first entry)

DE Tumour rejection antigen precursor coding sequence.

KW Tumour rejection antigen precursor; TRAP; TRA; melanoma; cancer;

KW tumour; treatment; detection; vaccine; HLA-A2; adoptive transfer;

KW T cell; T lymphocyte; human leukocyte antigen; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT misc\_feature 2685

FT /\*tag= a /note= "Unidentified nucleotide."

FT 9422..9456

FT misc\_feature /\*tag= b

FT /note= "This region is 4.7-5.3 kilobases in length

FT but the sequence of this region has not been

FT deduced."

FT 11539

FT misc\_feature /\*tag= c

FT /note= "Unidentified nucleotide."

FT WO9601557-A1.

PN 25-JAN-1996.

PD 27-JUN-1995; U08153.

PF 08-JUL-1994; US-272351.

PR 10-JAN-1995; US-370319.

PI (LUDWIG) LUDWIG INST CANCER RES.

PI Boon-falleur T, Brichard V, Coullie P, De PLAEN E;

PI Traversari C, Van PEL A, Wolfel T;

PI WPI: 96-097390/10.

DR Genomic DNA encoding a tumour rejection antigen precursor -

PT processed to antigen presented by HLA-A2, useful for treating or

PT diagnosing melanoma

PS Claim 3; Page 18-22; 41pp; English.

CC The tumour rejection antigen precursor (TRAP) is processed to a  
 CC tumour rejection antigen (TRA) presented by HLA-A2 molecules. TRA  
 CC is used to generate cytotoxic T lymphocytes for treating cancer (esp.  
 CC melanoma). It can also be used to raise specific antibodies, and  
 CC when complexed with HLA-A2, it can be used to produce vaccines.  
 CC Cytotoxic T lymphocytes so generated can be used in adoptive  
 CC transfer or generated, or they can be generated in vivo by using a  
 CC vector containing the appropriate gene or using TRA or TRAP together  
 CC with an adjuvant that facilitates entry into HLA-A2 presenting cells.  
 CC Diagnostic methods involving the detection of expression of TRAP  
 CC can be used in the detection of cancers.  
 SQ Sequence 13585 BP; 3827 A; 2859 C; 2968 G; 3894 T;

alignment\_scores:  
 Quality: 46.00 Length: 16  
 Ratio: 3.538 Gaps: 0  
 Percent Similarity: 81.250 Percent Identity: 56.250

alignment\_block:

US-08-653-294-35 x T11549 ..

Align seg 1/1 to: T11549 from: 1 to: 13585

4 LeuAsnArgLeuSerGluArgGluSerLeuArgGlnLeuArgGly 19  
 ::::|||||:|||||:|||||:|||||:||||| :||  
 246 ATCAACAGAGATTCTCCGCAACGTTTCTCAGTCTCCAACTCAGAGGG 293

OM of: US-08-653-294-35 to: EST:\* out\_format : pfs  
Date: Feb 8, 2000 6:23 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgnl\_1/USPTO.spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.2  
-DB=EST -QPMF=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500  
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-08-653-294-35  
Query length: 20  
Database: EST:\*  
Database sequences: 4538634  
Database length: 1887831982  
Search time (sec): 7600.090000

## score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_est11:AA319533	+	51.00	156.96	103	AA319533 EST70761 Adrenal gland
gb_est13:AA361477	+	51.00	149.93	189	AA361477 EST70761 T-cell lympho
gb_est13:AA352603	+	51.00	148.49	218	AA352603 EST60621 Activated T-c
gb_est13:AA329491	+	51.00	146.75	259	AA329491 EST100074 Pancreas tum
gb_est13:AA352960	+	51.00	146.33	270	AA352960 EST610074 Activated T-c
gb_est8:AA058454	+	51.00	143.06	373	AA058454 z167007.r1 Stratagene
gb_est11:AA024068	+	51.00	142.43	397	AA024068 z11102.r1 Stratagene
gb_est10:AA160317	+	51.00	142.23	405	AA160317 z056007.r1 Stratagene
gb_est9:AA100680	+	51.00	141.89	419	AA100680 z190802.r1 Stratagene
gb_est7:W40489	+	51.00	141.70	427	W40489 zc84b01.r1 Pancreatic is
gb_est12:AA310808	+	51.00	141.33	443	AA310808 EST181593 Jurkat T-cell
gb_est9:CI8310	+	51.00	140.62	475	CI8310 CI8310 Human placenta CD
gb_est17:AA663896	+	51.00	140.54	479	AA663896 ae74d01.s1 Stratagene
gb_est35:AL039796	+	51.00	140.04	503	AL039796 DKF29434B1912.r1 434
gb_est28:AT524732	+	51.00	138.55	583	AT524732 th12603.x1 NCI_CGAP_C
gb_est37:AA008453	+	51.00	137.30	660	AA008453 wv53005.x1 NCI_CGAP_Ga
gb_est11:AA263135	+	51.00	136.56	710	AA263135 PM0598 KGI-a Lambda 2
gb_gss3:B20346	-	51.00	135.31	804	B20346 F18J1-Sp6 IGF Arabidopsi
gb_gss10:AA0248102	-	50.50	137.61	67.21	AA0248102 HS-2015.A1.H09.MR CIT
gb_gss3:B90159	+	50.00	139.60	52.03	B90159 CIT-HSP-2173M11.TF CIT-H
gb_gss8:AA009853	+	50.00	138.93	392	AA009853 HS-2173.A1.G06.MF CIT
gb_gss3:B50728	-	50.00	136.69	523	B50728 CIT-HSP-623C13.TV CIT-HS
gb_gss13:AA0473632	+	50.00	134.64	641	AA0473632 CITBI-EI-2590L18.TF CIT
gb_gss13:AA0475628	+	50.00	133.50	717	AA0475628 CITBI-EI-2590L16.TF CI
gb_gss12:AA0377514	+	49.50	137.00	438	AA0377514 RPI-11-162p10.TJ RPI
gb_est5:D73898	-	49.00	140.37	47.14	D73898 CELK064C5F.Yu11 Kohara U
gb_est9:AA063950	+	49.00	138.57	395	AA063950 m143f02.r1 Stratagene
gb_gss11:AA029878	+	49.00	134.31	494	AA029878 HS-3005.B2.A11.R7 CIT
gb_est11:AA0205710	+	49.00	133.77	521	AA0205710 zq69a12.s1 Stratagene
gb_est34:AA1795124	+	48.50	142.17	196	AA1795124 sb7b04.y1 Gm-c1010 G1
gb_gss3:AA331579	+	48.00	138.77	37.91	AA331579 EST35512 Embryo, 8 wee
gb_est44:AW184423	-	48.00	132.31	449	AW184423 f114803.y1 Zebrafish a
gb_est35:AA1827894	-	48.00	132.05	461	AA1827894 w128006.s1 Soares_NFL
gb_est10:AA129456	-	48.00	131.56	484	AA129456 z178006.s1 Stratagene
gb_gss15:AA0641570	+	48.00	131.07	508	AA0641570 RPI-193-ECORI-319.TP RP
gb_est21:AA915979	+	48.00	130.83	520	AA915979 OH86605.s1 NCI_CGAP_C
gb_est11:AA216757	-	48.00	130.72	526	AA216757 zq96801.s1 Stratagene
gb_est44:AW187318	-	48.00	130.27	550	AW187318 BNHG112771 Six-day Cd
gb_est15:AA526987	-	48.00	130.16	556	AA526987 n119006.s1 NCI_CGAP_C
gb_gss11:AA029324	+	48.00	129.78	577	AA029324 HS-3111.A1.E09.MR CIT
gb_gss15:AA0641860	+	48.00	128.63	647	AA0641860 RPI-193-ECORI-6121.TV F
gb_est34:AA1794913	+	47.50	131.30	429	AA1794913 sb73d11.y1 Gm-c1010 G1

gb\_est34:AA1794877 + 47.50 129.22 197.12 527 ! A1794877 sb72h01.y1 Gm-cl010  
gb\_est34:AA1795011 + 47.50 129.22 197.12 527 ! A1795011 sb75a10.y1 Gm-cl010  
gb\_gss6:AA0873097 + 47.50 128.80 207.84 549 ! A0873097 V59C mTh-3xHA/lac2

seq\_name: gb\_est12:AA319533

## seq\_documentation\_block:

LOCUS AA319533 103 bp mRNA EST 19-APR-1997  
DEFINITION EST1772 Adrenal gland tumor Homo sapiens cDNA 5' end similar to  
similar to major histocompatibility complex, class I, B  
(GB:MI6102), mRNA sequence.

ACCESSION AA319533.1 GI:1971860  
VERSION AA319533  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 103)

REFERENCE

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wei,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
Shall,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseitine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.

Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL

MEDLINE

COMMENT On Sep 12, 1996 this sequence version replaced gi:1393672.

Other ESTs: THCI69519

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

1..103

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):120083"

/db\_xref="taxon:9606"

/clone\_lib="Adrenal gland tumor"

/dev\_stage="adult"

/note="Organ: Adrenal gland; Vector: pBluescript SK-;

Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 26 a 35 c 13 t

ORIGIN

alignment\_scores:

Quality: 51.00 Length: 10

Ratio: 5.100 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-35 x AA319533 ..

Align seg 1/1 to: AA319533 from: 1 to: 103

11 ArgGlusLeuArgAsnLeuArgGlyTyr 20  
|||||  
27 CGAGAGAGCCTCGCGAACCTTCGCGGCTAC 56

seq\_name: gb\_est13:AA361477

seq\_documentation\_block: 189 bp mRNA EST 21-APR-1997  
LOCUS AA361477 Homo sapiens cdna 5' end similar to  
DEFINITION EST70761 T-cell lymphoma Homo sapiens cdna 5' end similar to  
similar to major histocompatibility complex, class I, Bw62.3, mRNA  
sequence.

ACCESSION AA361477  
VERSION AA361477.1 GI:2013795  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 189)

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
Small,K.V., Springs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cdna sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
12140200  
On Sep 12, 1996 this sequence version replaced gi:1404737.  
Other ESTs: THC169519  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source

1..189  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):165623"  
/db\_xref="taxon:9606"  
/clone\_lib="T-cell lymphoma"  
/cell\_type="T-lymphocyte"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 45 a 60 c 55 g 25 t 4 others  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-35 x AA361477

Align seg 1/1 to: AA361477 from: 1 to: 189

11 ArgGlusLeuArgAsnLeuArgGlyTyr 20  
|||||  
37 CGAGAGAGCCTCGCGAACCTTCGCGGNTAC 66

seq\_name: gb\_est13:AA352603

seq\_documentation\_block: 218 bp mRNA EST 21-APR-1997  
LOCUS AA352603  
DEFINITION EST60621 Activated T-cells XX Homo sapiens cdna 5' end similar to  
similar to major histocompatibility complex, class I, Bw62.3, mRNA  
sequence.

ACCESSION AA352603  
VERSION AA352603.1 GI:2004923  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 218)

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
Small,K.V., Springs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cdna sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
12140200  
On May 8, 1995 this sequence version replaced gi:800964.  
Other ESTs: THC172938  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source

1..218  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):152802"  
/db\_xref="taxon:9606"  
/clone\_lib="Activated T-cells XX"  
/cell\_type="T-lymphocyte"  
/dev\_stage="adult"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 52 a 70 c 62 g 28 t 5 others  
ORIGIN



alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-35 x AA352603 ..  
 Align seg 1/1 to: AA352603 from: 1 to: 218

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
 |||||  
 39 CGAGAGAGCTGCGGAACCTGCGGGCTAC 68

seq\_name: gb\_est12:AA294911

seq\_documentation\_block: 259 bp mRNA EST 18-APR-1997  
 LOCUS AA294911  
 DEFINITION EST100074 Pancreas tumor I Homo sapiens cDNA 5' end similar to  
 similar to major histocompatibility complex, class I, Bw62.3, mRNA  
 sequence.

ACCESSION AA294911

VERSION AA294911.1 GI:1947266

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
 Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
 Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,  
 Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseitine,W.A., Fields,C.,  
 Fraser,C.M. and Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)

## JOURNAL

## MEDLINE

## COMMENT

On Nov 29, 1993 this sequence version replaced gi:430148.

Other ESTs: THCI72938

Contact: Kerlavage, AR

## Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

## FEATURES

## source

1..259

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):190413"

/db\_xref="taxon:9606"

/clone.lib="Pancreas tumor I"

/dev\_stage="adult"

/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI"

BASE COUNT 60 a 87 c 75 g 30 t 7 others

## ORIGIN

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-35 x AA294911 ..

Align seg 1/1 to: AA294911 from: 1 to: 259

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20

|||||

42 CGAGAGAGCTGCGGAACCTGCGGGTAC 71

seq\_name: gb\_est13:AA352960

seq\_documentation\_block: 270 bp mRNA EST 21-APR-1997

LOCUS AA352960

DEFINITION EST61101 Activated T-cells XX Homo sapiens cDNA 5' end similar to

similar to major histocompatibility complex, class I, B61, mRNA

sequence.

ACCESSION AA352960

VERSION AA352960.1 GI:2005353

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 270)

REFERENCE

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
 Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
 Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,  
 Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseitine,W.A., Fields,C.,  
 Fraser,C.M. and Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

On May 8, 1995 this sequence version replaced gi:801269.

Other ESTs: THCI72938

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..270

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):153240"

/db\_xref="taxon:9606"

/clone.lib="Activated T-cells XX"

/cell\_type="T-lymphocyte"

```

/dev_stage="adult"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

```

```

BASE COUNT      62 a   80 c   88 g   35 t   5 others
ORIGIN

```

```

alignment_scores:
  Quality:      51.00      Length:      10
  Ratio:        5.100      Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:

```

```

US-08-653-294-35 x AA352960 ..

```

```

Align seg 1/1 to: AA352960 from: 1 to: 270

```

```

11 ArgGusLeuArgAsnLeuArgGlyTyr 20
|||||
182 CGAGAGAGCTTCGGAACCTTCGCGGCTAC 211

```

```

seq_name: gb_est8:AA058454

```

```

seq_documentation_block:
LOCUS      AA058454      373 bp      mRNA      EST      01-DEC-1996
DEFINITION      216707.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
IMAGE:509677 5' similar to gb:M24039.cds1 HLA CLASS I
HISTOCOMPATIBILITY ANTIGEN, B-18 B*1801 ALPHA (HUMAN));, mRNA
sequence.

```

```

ACCESSION      AA058454
VERSION        AA058454.1 GI:1551280
KEYWORDS       EST.
SOURCE         human.

```

```

ORGANISM       Homo sapiens

```

```

REFERENCE
AUTHORS        Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

```

```

ORGANISM       Homo sapiens

```

```

REFERENCE
AUTHORS        Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

```

```

TITLE          Generation and analysis of 280,000 human expressed sequence tags
JOURNAL         Genome Res. 6 (9), 807-828 (1996)
MEDLINE        97044478
COMMENT         On Sep 12, 1996 this sequence version replaced gi:1393357.

```

```

Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

```

```

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1592 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 268.
Location/Qualifiers
1..373
/organism="Homo sapiens"
/db_xref="GDB:3813009"
/db_xref="taxon:9606"
/clone="IMAGE:509677"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGG 3' -3' adaptor sequence: 5'
CTCGAGTGTGTGTGTGTGTGT 3'"

```

```

FEATURES
source

```

```

BASE COUNT      73 a   119 c   116 g   57 t   8 others
ORIGIN

```

```

alignment_scores:
  Quality:      51.00      Length:      10
  Ratio:        5.100      Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:

```

```

US-08-653-294-35 x AA058454 ..

```

```

Align seg 1/1 to: AA058454 from: 1 to: 373

```

```

11 ArgGusLeuArgAsnLeuArgGlyTyr 20
|||||
229 CGAGAGAGCTTCGGAACCTTCGCGGCTAC 258

```

```

seq_name: gb_est11:AA224068

```

```

seq_documentation_block:
LOCUS      AA224068      397 bp      mRNA      EST      19-FEB-1997
DEFINITION      zrlif02.r1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone
IMAGE:648507 5' similar to gb:I22649.cds1 HLA CLASS I
HISTOCOMPATIBILITY ANTIGEN, B-39 B*3901 ALPHA (HUMAN));, mRNA
sequence.

```

```

ACCESSION      AA224068
VERSION        AA224068.1 GI:1844610
KEYWORDS       EST.
SOURCE         human.

```

```

ORGANISM       Homo sapiens

```

```

REFERENCE
AUTHORS        Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

```

```

ORGANISM       Homo sapiens

```

```

REFERENCE
AUTHORS        Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

```

```

TITLE          Generation and analysis of 280,000 human expressed sequence tags
JOURNAL         Genome Res. 6 (9), 807-828 (1996)
MEDLINE        97044478
COMMENT         On Nov 29, 1993 this sequence version replaced gi:430429.

```

```

Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

```

```

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev1 Et from Amersham
High quality sequence stop: 295.
Location/Qualifiers
1..397
/organism="Homo sapiens"
/db_xref="GDB:5583245"
/db_xref="taxon:9606"
/clone="IMAGE:648507"
/clone_lib="Stratagene hnt neuron (#937233)"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt.
Differentiated, post mitotic hnt neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGG 3' -3' adaptor sequence: 5'
CTCGAGTGTGTGTGTGTGT 3'"

```

```

FEATURES
source

```

```

BASE COUNT      77 a   129 c   131 g   56 t   4 others
ORIGIN

```

alignment\_scores:  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-35 x AA224068 ..

Align seg 1/1 to: AA224068 from: 1 to: 397

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
 |||||||  
 297 CGAGAGCGCTCGNAAACCTGCGCGCTAC 326

seq\_name: gb\_est10:AA160317

seq\_documentation\_block:  
 LOCUS AA160317 405 bp mRNA EST 09-MAR-1998  
 DEFINITION z056c07.t1 Stratagene pancreas (#937208) Homo sapiens cDNA clone  
 IMAGE:590892 5' similar to gb:U11265 HLA CLASS I HISTOCOMPATIBILITY  
 ANTIGEN, B-35 B\*3501 ALPHA (HUMAN);, mRNA sequence.

ACCESSION AA160317.1 GI:1734956

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;

## Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 405)

## Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

## Katzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,

## Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

## Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

## WashU-NCI human EST project

## Unpublished (1997)

## On Sep 12, 1996 this sequence version replaced gi:1406818.

## Contact: Wilson RK

## Washington University School of Medicine

## 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

## Tel: 314 286 1800

## Fax: 314 286 1810

## Email: est@watson.wustl.edu

## This clone is available royalty-free through LLNL; contact the

## IMAGE Consortium (info@image.llnl.gov) for further information.

## Insert Length: 1671 Std Error: 0.00

## Seq primer: -28M13 rev2 from Amersham

## High quality sequence stop: 332.

## Location/Qualifiers

## 1. 405

## /organism="Homo sapiens"

## /db\_xref="GDB:4622278"

## /db\_xref="taxon:9606"

## /clone="IMAGE:590892"

## /lab\_host="Stratagene pancreas (#937208)"

## /note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:

## EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:

## Oligo dt. Pancreatic adenocarcinoma cell line. Average

## insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

## sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'

## CTCAGCTTTTCTTTTCTTTT 3'

## BASE COUNT 80 a 132 c 132 g 58 t

## ORIGIN

## alignment\_scores:

Quality: 51.00 Length: 10

Ratio: 5.100 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-35 x AA160317 ..

Align seg 1/1 to: AA160317 from: 1 to: 405

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
 |||||||  
 281 CGAGAGCGCTCGGAACCTGCGCGCTAC 310

seq\_name: gb\_est9:AA100680

seq\_documentation\_block:

LOCUS AA100680 419 bp mRNA EST 31-JUL-1997

DEFINITION z190b02.r1 Stratagene colon (#937204) Homo sapiens cDNA clone

IMAGE:511851 5' similar to gb:M24039\_cds1 HLA CLASS I

HISTOCOMPATIBILITY ANTIGEN, B-18 B\*1801 ALPHA (HUMAN);, mRNA

sequence.

ACCESSION AA100680

VERSION AA100680.1 GI:1646981

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 419)

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1534 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 275.

Location/Qualifiers

1. 419

/organism="Homo sapiens"

/db\_xref="GDB:3844346"

/db\_xref="taxon:9606"

/clone="IMAGE:511851"

/clone\_lib="Stratagene colon (#937204)"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: colon; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. T-84 colonic epithelial cell line. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'

CTCAGCTTTTCTTTTCTTTT 3"

BASE COUNT 85 a 133 c 133 g 63 t

ORIGIN

alignment\_scores:

Quality: 51.00 Length: 16

Ratio: 3.643 Gaps: 1

Percent Similarity: 87.500 Percent Identity: 75.000

alignment\_block:

US-08-653-294-35 x AA100680 ..

Align seg 1/1 to: AA100680 from: 1 to: 419

5 AsnArgLeuSerGluArgGluSerLeuArgAsnLeuArgGlyTyr 20

```

225 CACAGACTTACCGAGAGA.....AGCTGCGGAACCTGCGCGCTAC 266
seq_name: gb_est7:W40489

seq_documentation_block:
LOCUS      W40489          427 bp      mRNA          EST          20-MAY-1996
DEFINITION zc84b01.r1 Pancreatic Islet Homo sapiens cDNA clone IMAGE:328969 5'
            similar to gb:U11265 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35
            B*3501 ALPHA (HUMAN);, mRNA sequence.
ACCESSION  W40489
VERSION    W40489.1  GI:1324496
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 427)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevisan,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE     Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   9704478
COMMENT   On Jan 25, 1995 this sequence version replaced gi:637865.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA-ET
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source           1..427
                     /organism="Homo sapiens"
                     /db_xref="GDB:1263173"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:328969"
                     /clone_lib="Pancreatic Islet"
                     /tissue_type="pancreatic islet"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)
                     Takeda et al. Cloned unidirectionally. Primer: Oligo dt.
                     -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTCTTTTCTTTT 3"
BASE COUNT      87 a 141 c 134 g 62 t 3 others
ORIGIN
alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-08-653-294-35 x W40489
Align seg 1/1 to: W40489 from: 1 to: 427
11 ArgGlusLeuArgAsnLeuArgGlyTyr 20
|||||
279 CGAGAGAGCTCGGAACCTGCGCGCTAC 308
seq_name: gb_est12:AA310808

```

```

seq_documentation_block:
LOCUS      AA310808          443 bp      mRNA          EST          19-APR-1997
DEFINITION EST181593 Jurkat T-cells V Homo sapiens cDNA 5' end similar to
            similar to major histocompatibility complex, class I, B0704, mRNA
            sequence.
ACCESSION  AA310808
VERSION    AA310808.1  GI:1963136
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 443)
AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Buit,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wal,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palances,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dinke,D., Feng,D.-F., Fertie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Korzak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE   12140200
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1397854.
            Other ESTs: THC180721
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlavet@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tadb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source           1..443
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):156811"
                     /db_xref="taxon:9606"
                     /clone_lib="Jurkat T-cells V"
                     /cell_type="T-lymphocyte"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      86 a 144 c 143 g 64 t 6 others
ORIGIN
alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-08-653-294-35 x AA310808
Align seg 1/1 to: AA310808 from: 1 to: 443
11 ArgGlusLeuArgAsnLeuArgGlyTyr 20
|||||
292 CGAGAGAGCTTCGGAACCTGCGCGCTAC 321

```

JOURNAL  
COMMENT

Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1394858.  
Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: polyr not found  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 430.

## FEATURES

Location/Qualifiers

1. 479  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:969889"  
/clone\_lib="Stratagene schizo brain S11"  
/sex="male"  
/tissue\_type="schizophrenic brain S-11 frontal lobe"  
/dev\_stage="34 years old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site: 1: EcoRI; Library  
constructed from S-11 frontal lobe, male, 34 years old,  
50% caucasian, 50% Aleutian. Schizophrenic suicide.  
Random primed into EcoRI site of ZAP II Vector. Mass  
excised. Custom library. Avg insert length 1.4Kb.  
Material obtained by Johnston N., Torrey, E.F., Yolken R.,  
and the Stanley Neuropathology Consortium - Analysis of  
RNAs from the Brains of Individuals with Psychiatric  
Diseases (Unpublished) Stanley Neurology Laboratory,  
Johns Hopkins School of Medicine, Baltimore MD."  
BASE COUNT 91 a 165 c 158 g 64 t 1 others  
ORIGIN

## alignment\_scores:

Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-35 x AA663896 ..

Align seg 1/1 to: AA663896 from: 1 to: 479

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
283 CGAGAGAGCCTGCGGAACCTGCGGCTAC 312

seq\_name: gb\_est35:AL039796

## seq\_documentation\_block:

LOCUS AL039796 503 bp mRNA EST 29-SEP-1999  
DEFINITION DKEZP43481912.r1 434 (synonym: htes3) Homo sapiens CDNA clone  
IMAGE:969889 3', similar to gb:M28203 HLA CLASS I HISTOCOMPATIBILITY  
ANTIGEN, B-62 B\*1504 ALPHA (HUMAN);, mRNA sequence.  
ACCESSION AL039796  
VERSION AL039796.1 GI:5408804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 503)  
AUTHORS Duesterhoeft,A., Lauber,J., Newes,H.W., Gassenhuber,J. and  
Wiemann,S.

TITLE EST (Duesterhoeft, et al.)

JOURNAL Unpublished (1999)

COMMENT On Jun 2, 1999 this sequence version replaced gi:4967270.

Contact: Duesterhoeft A

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

seq\_name: gb\_est9:C18310

seq\_documentation\_block: 475 bp mRNA EST 02-OCT-1996  
LOCUS C18310 Human placenta cDNA (TFujiwara) Homo sapiens CDNA clone  
DEFINITION GEN-560D07 5', mRNA sequence.

ACCESSION C18310.1 GI:1579912

VERSION C18310.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 475)

AUTHORS Fujiwara,T., Hirano,H., Hishigaki,H., Horie,M., Kawai,A., Kuga,Y.,  
Kyushiki,H., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y.,  
Shinomoya,H., Suzuki,M., Takaichi,A., Takeda,S., Watanabe,T.,  
Maekawa,H., Nakamura,Y. and Takahashi,E.

TITLE Otsuka cDNA project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1393837.

Contact: Tsutomu Fujiwara

Otsuka GEN Research Institute

Otsuka Pharmaceutical Co., Ltd

463-10 Kagasuno Kawauchi-cho, Tokushima, 771-01 Japan

Tel: 0886-65-2888

Fax: 0886-37-1035.

Location/Qualifiers

1. 475

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="GEN-560D07"

/clone\_lib="Human placenta cDNA (TFujiwara)"

/tissue\_type="placenta"

BASE COUNT 89 a 161 c 154 g 68 t 3 others  
ORIGIN

## alignment\_scores:

Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-35 x C18310 ..

Align seg 1/1 to: C18310 from: 1 to: 475

11 ArgGluSerLeuArgAsnLeuArgGlyTyr 20  
|||||  
299 CGAGAGAGCCTGCGGAACCTGCGGCTAC 328

seq\_name: gb\_est17:AA663896

## seq\_documentation\_block:

LOCUS AA663896 479 bp mRNA EST 15-DEC-1997  
DEFINITION ae74d01.sl Stratagene schizo brain S11 Homo sapiens CDNA clone  
IMAGE:969889 3', similar to gb:M28203 HLA CLASS I HISTOCOMPATIBILITY  
ANTIGEN, B-62 B\*1504 ALPHA (HUMAN);, mRNA sequence.

ACCESSION AA663896

VERSION AA663896.1 GI:2617887

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 479)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST project



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 04:05:44 ; Search time 133.56 Seconds  
(without alignments)  
2.128 Million cell updates/sec

Title: US-08-653-294-36  
Perfect score: 58  
Sequence: 1 YRLAIRRIALRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	12	1 R95429	HLA-B2702 84-79-84
2	58	100.0	12	1 W33798	Peptide B2702.84-7
3	58	100.0	12	1 W33799	Immunomodulating d
4	44	75.9	20	1 R92907	HLA-B2702 CTL modu
5	44	75.9	20	1 R95428	HLA-B2702 84-75-84
6	44	75.9	20	1 W33778	Immunomodulating d
7	39	67.2	20	1 R92909	HLA-B2702 CTL modu
8	39	67.2	20	1 R92908	HLA-B2702 CTL modu
9	39	67.2	20	1 W33791	Peptide B2702.84-7
10	39	67.2	20	1 W33792	Peptide B2702.84-7
11	37	63.8	20	1 R95430	HLA-B2702 84-75T/7
12	35	60.3	18	1 R71429	Human MHC 1 alpha
13	35	60.3	50	1 W44156	Human neuronal nic
14	34.5	59.5	20	1 R92911	HLA-B2702 CTL modu
15	34.5	59.5	20	1 W33779	Immunomodulating d
16	34	58.6	20	1 R92910	HLA-B2702 CTL modu
17	34	58.6	20	1 W33793	Peptide B2702.84-7
18	33	56.9	904	1 W86351	Human DNAX toll-11
19	32	55.2	40	1 W12037	Curvularia verrucu
20	32	55.2	600	1 W12042	Curvularia verrucu
21	32	55.2	627	1 W44152	Human neuronal nic
22	32	55.2	640	1 R82249	Chloroperoxidase.
23	31	53.4	111	1 R81449	Hepatitis GB virus
24	31	53.4	371	1 W31535	Xanthosine N7 meth
25	31	53.4	371	1 W65768	Coffee xanthosine
26	31	53.4	371	1 W76641	Coffee XMT protein
27	31	53.4	444	1 W41940	Corn p-hydroxyphen
28	31	53.4	504	1 W09022	Neuronal nicotinic
29	31	53.4	520	1 W74762	Human secreted pro
30	31	53.4	844	1 W69561	Human ligase IV. M
31	31	53.4	911	1 R95634	DNA-ligase-III. Ne
32	31	53.4	3567	1 R44431	eryA region polype
33	30	51.7	107	1 W26581	Human cytostatin I
34	30	51.7	107	1 W27561	Human cytostatin I

35	30	51.7	107	1	W83929	Human growth inhib
36	30	51.7	135	1	W30891	Human cytostatin I
37	30	51.7	347	1	W72010	HSV-2 strain SB5 C
38	30	51.7	450	1	W72145	HSV-2 strain SB5 C
39	30	51.7	546	1	W92336	M. rosa truncated
40	30	51.7	548	1	W72165	HSV-2 strain SB5 C
41	30	51.7	564	1	W92335	M. rosa fourth adh
42	30	51.7	2625	1	W55887	Human telomerase.
43	29	50.0	6	1	W47263	Immunomodulatory p
44	29	50.0	6	1	W47261	Immunomodulatory p
45	29	50.0	6	1	W33780	Peptide #1 used in

#### ALIGNMENTS

##### RESULT 1

R95429 R95429 standard; peptide; 12 AA.  
ID AC R95429;  
DT 12-NOV-1996 (first entry)  
DE HLA-B2702 84-79-84 palindrome.  
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN W0951328-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C. Krensky AM;  
DR WPI; 95-194027/25.  
PT Compsns. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12; 29pp; English.  
CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. This sequence represents the  
CC HLA-B2702 84-79-84 palindrome. These sequences can be used to isolate  
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane  
CC protein associated with T-cell activation in mammalian T-cells, and is  
CC also immunologically cross reactive with the heat shock protein Hsc70.  
CC p74 is found in a limited number of cell types, but is particularly  
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
CC cell with an amphoteric detergent, and then passed through an affinity  
CC column containing a covalently bound HLA-B2702 palindromic peptide.  
CC Compositions comprising the extracellular fragment of p74 combined with  
CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
CC compounds can be screened for their effect on the cytolytic activity of  
CC T-cells, by combining them with the extracellular portion of p74 and  
CC determining the amount of binding between the candidate compound and p74.  
CC Modulation of CTL activity can be inhibited in a cellular composition  
CC containing T-cells and antigen presenting cells (APCs), by adding to the  
CC mix the extracellular portion of p74, in an amount sufficient to compete  
CC with p74 for the binding of the p74 ligand.  
SQ Sequence 12 AA;

Query Match 100.0%; Score 58; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 3.3e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	YRLAIRRIALRY	12
DB	1	YRLAIRRIALRY	12

##### RESULT 2

W33798 W33798 standard; peptide; 12 AA.  
ID AC W33798;

DT 19-JUN-1998 (first entry)  
 DE Peptide B2702.84-79/79-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09744351-A1.  
 PD 27-NOV-1997: U086899.  
 PF 22-MAY-1997: US-653294.  
 PR 24-MAY-1996: US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DI WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W3784-98 and W3778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 58; DB 1; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12  
 |||||  
 Db 1 YRLAIRRIALRY 12

## RESULT 3

W33799 ID W33799 standard; peptide; 12 AA.  
 AC W33799;  
 DT 19-JUN-1998 (first entry)  
 DE Immunomodulating dimer peptide #3.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997: U086899.  
 PR 24-MAY-1996: US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DI WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Claim 17; Page 35; 41pp; English.  
 CC This sequence represents a specifically claimed immunomodulating  
 CC dimer peptide of the invention. A peptide-type compound or variant is

CC claimed which has immunomodulating activity, including the N-terminal  
 CC acylated and/or C-terminal amidated or esterified forms of up to 60  
 CC amino acids, where the peptide-type compound comprises the formula: A-B,  
 CC where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or  
 CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
 CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa  
 CC represents amino acid. The sequence in the brackets may optionally be  
 CC absent or truncated at any peptide type bond within the brackets. The  
 CC compounds comprise amino acid sequences related to a Class I HLA-B  
 CC alpha domain (positions 79-84). They can be used to inhibit cytotoxic  
 CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in  
 CC vitro. They can also be used in combination with antigenic peptides or  
 CC proteins of interest to activate CTLs. They can also inhibit the  
 CC proliferation of T cells in response to anti-CD3. The peptide can be  
 CC used for preventing rejection of transplants or for treating autoimmune  
 CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.  
 CC The products can also be used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 58; DB 1; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12  
 |||||  
 Db 1 YRLAIRRIALRY 12

## RESULT 4

R92907 ID R92907 standard; peptide; 20 AA.  
 AC R92907;  
 DT 16-MAY-1996 (first entry)  
 DE HLA-B\*2702 CTL modulating peptide (B2702.84-75/75-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW class I MHC; HLA-B\*2702.  
 OS Synthetic.  
 PN W09526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DI WPI: 95-358582/45.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B\*75-84 MHC antigen of the recipient  
 PT host  
 PS Example 15; Page 36; 80pp; English.  
 CC R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
 CC I MHC HLA-B\*2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with a  
 CC subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 SQ Sequence 20 AA;

Query Match 75.9%; Score 44; DB 1; Length 20;  
 Best Local Similarity 60.0%; Pred. No. 0.024;  
 Matches 12; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 YRLAIR-----RIALRY 12  
 |||||  
 Db 1 YRLAIRLNERRENRALRY 20  
 |||||



```

RESULT 5
R95428 ID R95428 standard; peptide; 20 AA.
AC R95428: 12-NOV-1996 (first entry)
DE HLA-B*2702 84-75-84 palindromic.
KW HLA: p74; alpha-helix; human-leucocyte-associated antigen; inhibitor;
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW cytolysis; antigen presenting cell.
OS Synthetic.
PN WO9513288-A1.
PD 18-NOV-1995.
PF 10-NOV-1994; US-150493.
PR 10-NOV-1993; US-150493.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
PT WPI; 95-194027/25.
DR Compsns. comprising lymphoid surface membrane proteins - which may
PT inhibit cytolytic activity and differentiation of CTLs.
PS Example; Page 12; 28pp; English.
CC R95413, and R95415-R95431 represent palindromes and fragments of
CC human-leucocyte-associated antigens. This sequence represents the
CC HLA-B*2702 84-75-84 palindromic. These sequences can be used to isolate
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC protein associated with T-cell activation in mammalian T-cells, and is
CC also immunologically cross reactive with the heat shock protein Hsc70.
CC p74 is found in a limited number of cell types, but is particularly
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC cell with an amphoteric detergent, and then passed through an affinity
CC column containing a covalently bound HLA-B*2702 palindromic peptide.
CC Compositions comprising the extracellular fragment of p74 combined with
CC HLA-B*2702 60-84 (see R95415), induces calcium influx, and inhibits
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC compounds can be screened for their effect on the cytolytic activity of
CC T-cells, by combining them with the extracellular portion of p74 and
CC determining the amount of binding between the candidate compound and p74.
CC Modulation of CTL activity can be inhibited in a cellular composition
CC containing T-cells and antigen presenting cells (APCs), by adding to the
CC mix the extracellular portion of p74, in an amount sufficient to compete
CC with p74 for the binding of the p74 ligand.
SQ Sequence 20 AA;

Query Match 75.9%; Score 44; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.024;
Matches 12; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 YRLAIR-----RIALRY 12
Db 1 YRLAIRLNERENRIALRY 20

RESULT 6
W33778 ID W33778 standard; peptide; 20 AA.
AC W33778: 19-JUN-1998 (first entry)
DE Immunomodulating dimer peptide #1.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
PN WO9744351-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
alpha-1 domain, used for preventing rejection of transplants or

```

```

PT treating autoimmune diseases
PS Claim 16; Page 35; 41pp; English.
CC This sequence represents a specifically claimed immunomodulating
CC dimer peptide of the invention. A peptide-type compound or variant is
CC claimed which has immunomodulating activity, including the N-terminal
CC acylated and/or C-terminal amidated or esterified forms of up to 60
CC amino acids, where the peptide-type compound comprises the formula: A-B,
CC where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E Or
CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a
CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa
CC represents amino acid. The sequence in the brackets may optionally be
CC absent or truncated at any peptide type bond within the brackets. The
CC compounds comprise amino acid sequences related to a Class I HLA-B
CC alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC vitro. They can also be used in combination with antigenic peptides or
CC proteins of interest to activate CTLs. They can also inhibit the
CC proliferation of T cells in response to anti-CD3. The peptide can be
CC used for preventing rejection of transplants or for treating autoimmune
CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.
CC The products can also be used for detection and diagnosis.
SQ Sequence 20 AA;

```

```

Query Match 75.9%; Score 44; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.024;
Matches 12; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

```

```

Qy 1 YRLAIR-----RIALRY 12
Db 1 YRLAIRLNERENRIALRY 20

```

```

RESULT 7
R92909 ID R92909 standard; peptide; 20 AA.
AC R92909: 16-MAY-1996 (first entry)
DE HLA-B*2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLA-B*2702.
OS Synthetic.
PN WO9526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349
PR 05-APR-1994; US-222851.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI; 95-358582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B75-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC I MHC HLA-B*2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

```

```

Query Match 67.2%; Score 39; DB 1; Length 20;
Best Local Similarity 55.0%; Pred. No. 0.21;
Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

```

```

Qy 1 YRLAIR-----RIALRY 12

```

Query Match 67.2%; Score 39; DB 1; Length 20;

Best Local Similarity 55.0%; Pred. No. 0.21;  
Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 YRLAIR-----RIALRY 12  
DB 1 YRLAIRLNRRNRLRALRY 20

RESULT 11  
R95430  
ID R95430 standard; peptide; 20 AA.  
AC R95430;  
DE HLA-B\*2702 84-75T/75-84T palindromic.  
KW HLA: p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN WO9513288-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM;  
DR WPI: 95-194027/25.  
PT Compsns. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12; 29pp; English.  
CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. This sequence represents the  
CC HLA-B\*2702 84-75T/75-84T palindromic. These sequences can be used to  
CC isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface  
CC membrane protein associated with T-cell activation in mammalian T-cells,  
CC and is also immunologically cross reactive with the heat shock protein  
CC Hsc70. p74 is found in a limited number of cell types, but is  
CC particularly expressed on B and T cells. p74 can be isolated by lysis of  
CC a suitable cell with an amphoteric detergent, and then passed through an  
CC affinity column containing a covalently bound HLA-B\*2702 palindromic  
CC peptide. Compositions comprising the extracellular fragment of p74  
CC combined with HLA-B\*2702.60-84 (see R95416), induces calcium influx, and  
CC inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis.  
CC Candidate compounds can be screened for their effect on the cytolytic  
CC activity of T-cells, by combining them with the extracellular portion of  
CC p74 and determining the amount of binding between the candidate compound  
CC and p74. Modulation of CTL activity can be inhibited in a cellular  
CC composition containing T-cells and antigen presenting cells (APCs), by  
CC adding to the mix the extracellular portion of p74, in an amount  
CC sufficient to compete with p74 for the binding of the p74 ligand.  
SQ Sequence 20 AA;

Query Match 63.8%; Score 37; DB 1; Length 20;  
Best Local Similarity 57.9%; Pred. No. 0.51;  
Matches 11; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 YRLAIR-----RIALR 11  
DB 1 YRLAIRLNRRNRLRALR 19

RESULT 12  
R71429  
ID R71429 standard; peptide; 18 AA.  
AC R71429;  
DE 12-OCT-1995 (first entry)  
DE Human MHC 1 alpha 1 domain peptide [Ala81]-Dk-(69-85).  
KW Major histocompatibility complex class 1; MHC 1; cell receptor;  
KW alpha 1 domain; peptide [Ala81]-Dk-(69-85); interaction modulation;  
KW arthritis; neoplasias; lupus erythematosus.  
OS Homo sapiens.  
PN WO9505189-A.  
PD 23-FEB-1995.

PF 12-AUG-1994; U09189.  
PR 12-AUG-1993; US-105416.  
PA (REGC) UNIV CALIFORNIA.  
PI Goldstein A, Goodenow RS, Olsson L;  
DR WPI: 95-098577/13.  
PT Regulating cell surface receptor response - by modulating  
PT interaction between MHC class I antigen and the cell surface  
PT receptor  
PS Example 4; Page 45; 103pp; English.  
CC R71424-R71438 are human major histocompatibility complex class 1  
CC (MHC 1) alpha 1 domain derived peptides and peptide mutants, they  
CC were used to modulate interactions between MHC 1 and cell surface  
CC receptors. Via competitive inhibition the peptides diminish the  
CC receptors response, this feature may be useful for the treatment  
CC of neoplasias, lupus erythematosus and arthritis.  
SQ Sequence 18 AA;

Query Match 60.3%; Score 35; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 1.1;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12  
DB 6 FRVDELTALRY 17

RESULT 13  
W44156  
ID W44156 standard; Protein; 504 AA.  
AC W44156;  
DE 14-MAY-1998 (first entry)  
DE Human neuronal nicotinic acetylcholine receptor alpha-3 subunit.  
KW Human; neuronal nicotinic acetylcholine receptor; alpha-3 subunit;  
KW brain tissue; screening; NACHR; antibody.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..30  
FT /label= signal  
FT Misc\_difference 235  
FT /label= unspecified  
FT /note= "encoded by WWC"  
FT Domain 240..265  
FT /label= TMD1  
FT /note= "transmembrane domain"  
FT Misc\_difference 246  
FT /label= unspecified  
FT /note= "encoded by AWC"  
FT Domain 273..296  
FT /label= TMD2  
FT /note= "transmembrane domain"  
FT Misc\_difference 275  
FT /note= "encoded by TGY"  
FT Domain 302..326  
FT /label= TMD3  
FT /note= "transmembrane domain"  
FT Misc\_difference 309  
FT /label= unspecified  
FT /note= "encoded by WWC"  
FT Misc\_difference 347  
FT /label= unspecified  
FT /note= "encoded by AMC"  
FT Misc\_difference 354  
FT /label= unspecified  
FT /note= "encoded by TWC"  
FT Domain 459..480  
FT /label= TMD4  
FT /note= "transmembrane domain"  
FT Region 327..458  
FT /label= cytoplasmic\_loop  
PN WO9420617-A2.  
PD 15-SEP-1994.  
PF 08-MAR-1994; U02447.

PR 08-MAR-1993; US-028031.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PA (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Elliott KJ, Ellis SB, Harpold MM;  
 DR WPI; 94-303024/37.  
 DR N-PSDB; V12200.  
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
 PT also transformed cells useful for screening cpds. which modulate  
 PT activity of the receptor  
 PS Example 2; Page 72-73; 99pp; English.  
 CC The present sequence represents a human neuronal nicotinic acetylcholine  
 CC receptor (NAChR) subunit. The cells expressing the alpha and/or beta  
 CC NAChR subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal NAChR.  
 CC Subunit specific antibodies may be used to monitor the distribution  
 CC and expression density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits or specific receptor  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes.  
 SQ Sequence 504 AA;

Query Match 60.38; Score 35; DB 1; Length 504;  
 Best Local Similarity 56.3%; Pred. No. 32;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLAIRRLRY 12  
 I | | | | |  
 Db 232 YSLIRLSLEY 243

RESULT 14  
 R92911  
 ID R92911 standard; peptide; 20 AA.  
 AC R92911; 1996 (first entry)  
 DT 16-MAY-1996  
 DE HLA-B\*2702 CTL modulating peptide (B2702.84-75/84-75).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW class I MHC; HLA-B\*2702.  
 OS Synthetic.  
 PN WO9526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI; 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B\*7-84 MHC antigen of the recipient  
 PT host  
 PS Example 15; Page 36; 80pp; English.  
 CC R3061-R33085, R83090-R83096 and R92907-R92914 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC is a dimer of residues 75-84 of the alpha-1 domain of the class I MHC  
 CC HLA-B\*2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with a  
 CC subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 SQ Sequence 20 AA;

Query Match 59.5%; Score 34.5; DB 1; Length 20;  
 Best Local Similarity 56.2%; Pred. No. 1.5;

Matches 9; Conservative 2; Mismatches 0; Indels 5; Gaps 1;  
 QY 1 YRLAIR-----RIALR 11  
 | | | | | | | | | |  
 Db 1 YRLAIRLNERYRLAIR 16

## RESULT 15

W33779  
 ID W33779 standard; peptide; 20 AA.  
 AC W33779;  
 DT 19-JUN-1998 (first entry)  
 DE Immunomodulating dimer peptide #2.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI; 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PT Claim 16; Page 35; 41pp; English.  
 CC This sequence represents a specifically claimed immunomodulating  
 CC dimer peptide of the invention. A peptide-type compound or variant is  
 CC claimed which has immunomodulating activity, including the N-terminal  
 CC acylated and/or C-terminal amidated or esterified forms of up to 60  
 CC amino acids, where the peptide-type compound comprises the formula: A-B,  
 CC where A, B = (R aa76-77L) or (aa84-79) or (aa84-79) (Laa77-76R); aa76 = E or  
 CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
 CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa  
 CC represents amino acid. The sequence in the brackets may optionally be  
 CC absent or truncated at any peptide type bond within the brackets. The  
 CC compounds comprise amino acid sequences related to a Class I HLA-B  
 CC alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic  
 CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in  
 CC vitro. They can also be used in combination with antigenic peptides or  
 CC proteins of interest to activate CTLs. They can also inhibit the  
 CC proliferation of T cells in response to anti-CD3. The peptide can be  
 CC used for preventing rejection of transplants or for treating autoimmune  
 CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.  
 CC The products can also be used for detection and diagnosis.  
 SQ Sequence 20 AA;

Query Match 59.5%; Score 34.5; DB 1; Length 20;  
 Best Local Similarity 56.2%; Pred. No. 1.5;

Matches 9; Conservative 2; Mismatches 0; Indels 5; Gaps 1;

QY 1 YRLAIR-----RIALR 11  
 | | | | | | | | | |  
 Db 1 YRLAIRLNERYRLAIR 16

Search completed: February 8, 2000, 04:05:45  
 Job time: 9362 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 18:04:41 ; Search time 111.22 Seconds  
(without alignments)  
5.089 Million cell updates/sec

Title: US-08-653-294-36  
Perfect score: 58  
Sequence: 1 YRLAIRRALRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR\_62.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	69.0	309	2 G70882	probable oxidoredu
2	36	62.1	313	2 JC5342	Na+/H+ antiporter
3	36	62.1	348	2 S29990	histocompatibility
4	35	60.3	276	2 S74745	ABC-type transport
5	35	60.3	281	2 S44230	strf protein - Str
6	35	60.3	281	2 S17776	strf protein - Str
7	35	60.3	388	2 S15593	hypothetical prote
8	34	58.6	178	2 D72507	hypothetical prote
9	34	58.6	267	2 S44228	strf protein - Str
10	34	58.6	267	2 S44225	strf protein - Str
11	34	58.6	469	2 C69460	conserved hypothet
12	34	58.6	573	2 F72691	probable fumarate
13	34	58.6	4367	1 B54802	dyein heavy chain
14	33	56.9	128	2 S69745	hypothetical prote
15	33	56.9	141	2 H71504	ribosomal protein
16	33	56.9	190	2 F70410	hypothetical prote
17	33	56.9	264	2 S2102	thioether S-methyl
18	33	56.9	477	2 G69760	beta-glucosidase h
19	33	56.9	518	2 S72938	hflx protein - Myc
20	33	56.9	682	2 A41798	carboxy-terminal p
21	33	56.9	1124	2 D65032	hypothetical prote
22	32	55.2	142	2 C72053	L17 ribosomal prot
23	32	55.2	203	1 S50980	NHP10 protein - ye
24	32	55.2	215	2 T16542	hypothetical prote
25	32	55.2	252	2 I48120	P-glycoprotein - C
26	32	55.2	253	2 E75055	hypothetical prote
27	32	55.2	259	2 I48119	P-glycoprotein - C
28	32	55.2	306	2 F72323	conserved hypothet
29	32	55.2	339	2 D72509	hypothetical prote
30	32	55.2	431	1 BWBSSY	preprotein translo

31	32	55.2	434	2 T04960	hypothetical prote
32	32	55.2	609	2 S69334	chloride peroxidase
33	32	55.2	857	2 T05172	hypothetical prote
34	32	55.2	978	2 T00336	hypothetical prote
35	32	55.2	1010	2 T16616	hypothetical prote
36	32	55.2	1281	2 I48123	P-glycoprotein iso
37	32	55.2	1998	2 T08822	nonstructural poly
38	31	53.4	77	2 C70967	hypothetical prote
39	31	53.4	78	2 F64429	DNA-directed RNA p
40	31	53.4	153	2 S08501	3-dehydroquinatase
41	31	53.4	245	2 D70882	probable dapB prot
42	31	53.4	264	2 E71567	hypothetical prote
43	31	53.4	264	2 S76802	hypothetical prote
44	31	53.4	265	2 S74282	hypothetical prote
45	31	53.4	286	2 T02119	hypothetical prote

## ALIGNMENTS

## RESULT 1

G70882

Probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 29-Sep-1999

C:Accession: G70882

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Nature 393, 537-544, 1998A: Authors: Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Qua  
; Taylor, K.; Whitehead, S.; Barrell, B.G.A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A: Reference number: A70500; MUID: 98295987

A: Accession: G70882

A: Status: Preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-309 &lt;COL&gt;

A: Cross-references: GB: AL008967; GB: AL123456; NID: g3261491; PIDN: CAA15591.1; PID: g129

C: Geneticals:

A: Gene: RV2776C

C: Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; fer  
F: 9-211/Domain: cytochrome-b5 reductase homology <CBR>

F: 240-297/Domain: ferredoxin [2Fe-2S] homology &lt;FER&gt;

Query Match 69.0%; Score 40; DB 2; Length 309;  
Best Local Similarity 88.9%; Pred. No. 2;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YRLAIRRIA 9

||:|||||

Db 65 YRLAIRRIA 73

## RESULT 2

JC5342

Na+/H+ antiporter cdv2 - Clostridium difficile

C: Species: Clostridium difficile

C: Date: 27-May-1997 #sequence\_revision 18-Jul-1997 #text\_change 10-Sep-1997

C: Accession: JC5342

R: Braun, V.; Hundsberger, T.; Leukel, P.; Sauerborn, M.; von Eichel-Streiber, C.

Gene 181, 29-38, 1996

A: Title: Definition of the single integration site of the pathogenicity locus in Clos  
A: Reference number: JC5340; MUID: 97128764

A: Accession: JC5342

A: Molecule type: DNA

A: Residues: 1-313 &lt;CBRA&gt;

A: Cross-references: EMBL: X29282; NID: g1770128; PID: e212007; PID: g1770130

A: Experimental source: strain VP10463

C: Comment: This protein contains 11 membrane spanning domain. It functionally links t

C: Geneticals:

A: Gene: cdv2

Query Match 62.1%; Score 36; DB 2; Length 313;  
 Best Local Similarity 41.7%; Pred. No. 12;  
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12  
 ||: ||:||||  
 Db 259 YRVVARRLSIKY 270

RESULT 3  
 S29990  
 histocompatibility antigen, HLA-F-like - rhesus macaque  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
 C:Accession: S29990  
 R:Bontrop, R.R.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S29990  
 A:Accession: S29990  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-348 <BON>  
 A:Cross-references: EMBL:Z21819; NID:g38568; PIDN:CAA79885.1; PID:g38569  
 C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
 C:Keywords: transmembrane protein  
 F:219-284/Domain: immunoglobulin homology <IMM>

Query Match 62.1%; Score 36; DB 2; Length 348;  
 Best Local Similarity 54.5%; Pred. No. 14;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRRIALRY 12  
 ||: ||: |||  
 Db 98 RVALRKLLRY 108

RESULT 4  
 S74745  
 ABC-type transport protein slr0977 - Synechocystis sp. (strain PCC 6803)  
 N:Alternate names: protein slr0977  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
 C:Accession: S74745  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 S.  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S74745  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-276 <NAN>  
 A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017629; PID:g1651970  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Keywords: transport protein

Query Match 60.3%; Score 35; DB 2; Length 276;  
 Best Local Similarity 56.7%; Pred. No. 17;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12  
 ||: ||: |||  
 Db 31 YTLAWRDIAVRY 42

RESULT 5  
 S44230

stf protein - Streptomyces glaucescens  
 C:Species: Streptomyces glaucescens  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Sep-1997  
 C:Accession: S44230  
 R:Mayer, G.; Piepersberg, W.  
 submitted to the EMBL Data Library, April 1994  
 A:Reference number: S44224  
 A:Accession: S44230  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-281 <MAY>  
 A:Cross-references: EMBL:X78974; NID:g475235; PID:g581673  
 C:Genetics:  
 A:Start codon: GTG

Query Match 60.3%; Score 35; DB 2; Length 281;  
 Best Local Similarity 70.0%; Pred. No. 17;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRRIALR 11  
 ||: ||: |||  
 Db 234 RIAARRLALR 243

RESULT 6  
 S17776  
 stf protein - Streptomyces griseus  
 C:Species: Streptomyces griseus  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 20-Mar-1998  
 C:Accession: S17776  
 R:Mansouri, K.; Piepersberg, W.  
 Mol. Gen. Genet. 228, 459-469, 1991  
 A:Title: Genetics of streptomycin production in Streptomyces griseus: nucleotide sequ  
 A:Reference number: S17775; MUID:91375432  
 A:Accession: S17776  
 A>Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-281 <MAN>  
 A:Cross-references: GB:Y00459; GB:S55493; NID:g1621271; PID:e275192; PID:g1621272  
 C:Genetics:  
 A:Gene: stf  
 A:Start codon: GTG

Query Match 60.3%; Score 35; DB 2; Length 281;  
 Best Local Similarity 70.0%; Pred. No. 17;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRRIALR 11  
 ||: ||: |||  
 Db 234 RLAAARRLALR 243

RESULT 7  
 S15593  
 hypothetical protein (insertion sequence ISH27-3) - Halobacterium halobium  
 C:Species: Halobacterium halobium  
 C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 30-Jun-1998  
 C:Accession: S15593  
 R:Pfeifer, F.; Blaseio, U.  
 Nucleic Acids Res. 18, 6921-6925, 1990  
 A:Title: Transposition burst of the ISH27 insertion element family in Halobacterium h  
 A:Reference number: S15591; MUID:91088266  
 A:Accession: S15593  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-388 <PFE>  
 A:Cross-references: EMBL:X54434  
 A:Note: the authors translated the initiation codon GTG for residue 1 as Val  
 C:Genetics:  
 A:Mobile element: insertion sequence ISH27-3  
 A:Start codon: GTG

Query Match 60.3%; Score 35; DB 2; Length 388;

Best Local Similarity 75.0%; Pred. No. 23; Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLAIRRI 8

|||||:

Db 154 YRLAVRRL 161

RESULT 8

D72507 hypothetical protein APE2034 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: D72507

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339

A:Accession: D72507

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-178 <KAW>

A:Cross-references: DDBJ:AP000063; NID:g5105654; PIDN:BAA81044.1; PID:dl044830; PID:g510

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE2034

Query Match

Best Local Similarity 58.6%; Score 34; DB 2; Length 178;

Matches 7; Conservative 2; Mismatches 2; Indels 2; Indels 0; Gaps 0;

QY 2 RLAIRRIALRY 12

|||:|||||:

Db 34 RMACRRIALVY 44

RESULT 9

S44228

strf protein - Streptomyces galbus

C:Species: Streptomyces galbus

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Sep-1997

C:Accession: S44228

R:Mayer, G.; Piepersberg, W.

submitted to the EMBL Data Library, April 1994

A:Reference number: S44224

A:Accession: S44228

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-267 <MAY>

A:Cross-references: EMBL:X78973; NID:g475232; PID:g475234

Query Match

Best Local Similarity 58.6%; Score 34; DB 2; Length 267;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIRRIALR 11

|||:|||||:

Db 234 RLARRLAVR 243

RESULT 10

S44225

strf protein - Streptomyces bluenis (fragment)

C:Species: Streptomyces bluenis

C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 10-Sep-1997

C:Accession: S44225

R:Mayer, G.; Piepersberg, W.

submitted to the EMBL Data Library, April 1994

A:Reference number: S44224

A:Accession: S44225

A:Molecule type: DNA

A:Residues: 1-267 <MAY>

A:Cross-references: EMBL:X78972; NID:g475227; PID:g581604

C:Genetics:

A:Start codon: GTG

Query Match

Best Local Similarity 58.6%; Score 34; DB 2; Length 267;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIRRIALR 11

|||:|||||:

Db 234 RVARRLALR 243

RESULT 11

C69460

conserved hypothetical protein AF1684 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: C69460

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343

A:Accession: C69460

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-469 <KLE>

A:Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89565.1; PID:g264

C:Superfamily: hypothetical protein M30966

Query Match

Best Local Similarity 58.6%; Score 34; DB 2; Length 469;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12

|||:|||||:

Db 12 YRKGMKRIALVY 23

RESULT 12

F72691

probable fumarate reductase flavoprotein subunit APE0950 - Aeropyrum pernix (strain K

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: F72691

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: F72691

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-573 <KAW>

A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79934.1; PID:dl043720; PID:g

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0950

Query Match

Best Local Similarity 58.6%; Score 34; DB 2; Length 573;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 AIRRIALRY 12

:-:|:|:|

Db 320 AVREIAIRY 328

RESULT 13

B54802

dyein heavy chain, cytosolic - Neurospora crassa

N:Contains: dynein ATPase (EC 3.6.1.33)

C:Species: Neurospora crassa

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: B54802

R:Planmann, M.; Minke, P.F.; Tinsley, J.H.; Bruno, K.S.

J. Cell Biol. 127, 139-149, 1994

A:Title: Cytoplasmic dynein and actin-related protein Arp1 are required for normal nucle

A:Reference number: A54802; MUID:95014704

A:Accession: B54802

A:Molecule type: DNA

A:Residues: 1-4367 <PLA>

A:Cross-references: GB:L31504; NID:9473489; PID:9473490

C:Genetics:

A:Insertions: 104/1; 4205/3

C:Superfamily: dynein heavy chain, cytosolic

C:Keywords: ATP; heterotrimer; hydrolase; microtubule binding; P-loop

F:1943-1950/Region: nucleotide-binding motif A (P-loop)

F:2240-2247/Region: nucleotide-binding motif A (P-loop)

F:2603-2612/Region: nucleotide-binding motif A (P-loop)

F:2947-2954/Region: nucleotide-binding motif A (P-loop)

F:1949/Binding site: ATP (lys) #status predicted

F:2246/Binding site: ATP (lys) #status predicted

F:2611/Binding site: ATP (lys) #status predicted

F:2953/Binding site: ATP (lys) #status predicted

Query Match

Best Local Similarity 58.6%; Score 34; DB 1; Length 4367;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 AIRRIALRY 12

:-:|:|:|

Db 2859 AVRIAMEY 2867

RESULT 14

S69745

hypothetical protein YDR360w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C>Date: 23-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 26-Aug-1999

C:Accession: S69745

R:Du, Z.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of S. cerevisiae cosmid 9476.

A:Reference number: S61146

A:Accession: S69745

A:Molecule type: DNA

A:Residues: 1-128 <DUZ>

A:Cross-references: EMBL:U28372; NID:9849170; PID:92194159; MIPS:YDR360w

C:Genetics:

A:Map position: 4R

C:Superfamily: Saccharomyces hypothetical protein YDR360w

Query Match

Best Local Similarity 56.9%; Score 33; DB 2; Length 128;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLAIRRLR 11

:-:|:|:|

Db 113 YILSVRLRLR 123

RESULT 15

H71504

ribosomal protein L17 - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 13-Aug-1999

C:Accession: H71504; I40747

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t

A:Reference number: A71570; MUID:99000809

A:Accession: H71504

A:Molecule type: DNA

A:Residues: 1-141 <ARN>

A:Cross-references: GB:AE001273; NID:93328931; PIDN:AAC68107.1; PID:g332

A:Experimental source: serotype D, strain UW-3/Cx

R:Gu, L.; Wenman, W.M.; Ramacha, M.; Meuser, R.; Coffin, J.; Kaul, R.

J. Bacteriol. 177, 2594-2601, 1995

A:Title: Chlamydia trachomatis RNA polymerase alpha subunit: sequence and structural

A:Reference number: I40743; MUID:95247702

A:Accession: I40747

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 20-115, R, 117-141 <GUL>

A:Cross-references: GB:L33834; NID:9620026; PIDN:AAA74990.1; PID:9620030

C:Genetics:

A:Gene: rll7

C:Superfamily: Escherichia coli ribosomal protein L17

C:Keywords: protein biosynthesis; ribosome

Query Match

Best Local Similarity 56.9%; Score 33; DB 2; Length 141;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLAIRRIALRY 12

:-:|:|:|

Db 64 RLAVGLMVRV 74

Search completed: February 7, 2000, 18:04:42

Job time: 22208 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:26:02 ; Search time 68.63 seconds  
(without alignments)  
5.222 Million cell updates/sec

Title: US-08-653-294-36  
Perfect score: 58  
Sequence: 1 YRLAIRRIALRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	62.1	348	1 HLA_F_MACMU	P33617 macaca mula
2	35	60.3	281	1 STRF_STRGR	P03397 streptomyc
3	34	58.6	4367	1 DYHC_NEUCR	P45443 neurospora
4	33	56.9	122	1 RL17_CHLTR	P47760 chlamydia t
5	33	56.9	264	1 TSMT_MOUSE	P40936 mus musculu
6	33	56.9	477	1 BG12_BACSU	P42403 bacillus su
7	33	56.9	682	1 PRC_ECOLI	P23865 escherichia
8	33	56.9	1124	1 YPHG_ECOLI	P7585 escherichia
9	32	55.2	203	1 NH10_YEAST	Q03435 saccharomyc
10	32	55.2	375	1 HRMA_PSEY	Q08370 pseudomonas
11	32	55.2	431	1 SECY_BACSU	P16336 bacillus su
12	32	55.2	487	1 ACH6_CAEEL	Q93149 caenorhabdi
13	32	55.2	609	1 PRXC_CURIN	P49053 curvularia
14	32	55.2	723	1 YE7A_SCHPO	O14162 schizosacch
15	32	55.2	1281	1 MDR3_CRIGR	P23174 cricetus
16	31	53.4	78	1 RPOH_METVA	Q58443 methanococc
17	31	53.4	153	1 3DHQ_EMENI	P05147 emericella
18	31	53.4	212	1 YCD4_YEAST	P25369 saccharomyc
19	31	53.4	245	1 DAPB_MYCTU	O06601 mycobacteri
20	31	53.4	349	1 BIOB_MYCTU	P31137 escherichia
21	31	53.4	355	1 DEGS_ECOLI	P42381 coxiella bu
22	31	53.4	367	1 DNAJ_COXBU	Q07263 bos taurus
23	31	53.4	495	1 ACH3_BOVIN	P09481 gallus gall
24	31	53.4	496	1 ACH3_CHICK	P04757 rattus norv
25	31	53.4	499	1 ACH3_RAT	P32297 homo sapien
26	31	53.4	503	1 ACH3_HUMAN	Q57538 haemophilus
27	31	53.4	552	1 Y664_HAEIN	P34794 brassica na
28	31	53.4	582	1 RUB2_BRANA	P48537 cercospora
29	31	53.4	621	1 CRTI_CERNC	P45306 haemophilus
30	31	53.4	695	1 PRC_HAEIN	P09284 varicella-z
31	31	53.4	770	1 PRTP_VZVD	P49917 homo sapien
32	31	53.4	844	1 DNL4_HUMAN	P32843 saccharomyc
33	31	53.4	850	1 RN12_YEAST	Q91437 squalus aca
34	31	53.4	2242	1 PYR1_SQUAC	

35 31 53.4 3567 1 ERY2\_SACER Q03132 saccharopol  
36 31 53.4 4344 1 DYHC\_EMENI P45444 emericella  
37 30.5 52.6 491 1 HEMN\_ALCEU O34162 alcaligenes  
38 30 51.7 124 1 VJ05\_VACCV P07618 vaccinia v1  
39 30 51.7 133 1 VJ05\_VACCV P21083 vaccinia v1  
40 30 51.7 133 1 VJ05\_VARV P33055 variola vir  
41 30 51.7 145 1 TASM\_BFDV P13895 budgerigar  
42 30 51.7 185 1 PTH\_RICPR O95cv4 rickettsia  
43 30 51.7 226 1 HMC5\_DESVH P33392 desulfovibr  
44 30 51.7 310 1 DNL4\_MYGE P47248 mycoplasma  
45 30 51.7 323 1 VANH\_ENTFA Q47748 enterococcu

## ALIGNMENTS

RESULT 1  
HLAF\_MACMU  
ID HLAF\_MACMU STANDARD; PRT; 348 AA.  
AC P33617;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F ANTIGEN) (LEUKOCYTE ANTIGEN F).  
GN HLA-F OR HLAF.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 93246295.  
RA OTTING N., BONTROP R.E.;  
RT "Characterization of the rhesus macaque (Macaca mulatta) equivalent of HLA-F";  
RL Immunogenetics 38:141-145(1993).  
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z21819; CAA79885.1; -  
DR PIR: S29990; S29990.  
DR HSSP: P03989; IHSA.  
DR PROSITE: PS00290; IG\_MHC; 1.  
DR PFAM: PF00047; Ig; 1.  
DR PFAM: PF00129; MHC.I; 1.  
KW MHC I; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 348  
FT BY SIMILARITY.  
FT ALPHA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
FT ALPHA CHAIN F.  
FT EXTRACELLULAR ALPHA-1.  
FT EXTRACELLULAR ALPHA-2.  
FT EXTRACELLULAR ALPHA-3.  
FT CONNECTING PEPTIDE.  
FT CYTOPLASMIC TAIL.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT CARBOHYD POTENTIAL.  
SQ SEQUENCE 348 AA; 39300 MW; 3A375142 CRC32;

Query Match

62.1%; Score 36; DB 1; Length 348;

Best Local Similarity 54.5%; Pred. No. 5.2;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRIALRY 12  
|:|:|:|:|  
Db 98 RVALRKLLRY 108

## RESULT 2

STRE\_STRGR STANDARD; PRT; 281 AA.  
ID STRE\_STRGR  
AC P09397;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE STREPTOMYCIN BIOSYNTHESIS PROTEIN STRF.  
GN STRF.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2-3-11;  
RX MEDLINE; 91375432.  
RA MANSOURI K., PIEPERSBERG W.;  
RT "Genetics of streptomycin production in Streptomyces griseus:  
RT nucleotide sequence of five genes, strFGHIK, including a phosphatase  
RT gene.";  
RL Mol. Gen. Genet. 228:459-469(1991).  
CC -!- FUNCTION: MAY BE INVOLVED IN THE FORMATION OF N-METHYL-L-  
CC GLUCOSAMINE.  
CC -!- PATHWAY: STREPTOMYCIN BIOSYNTHESIS.  
CC -!- SIMILARITY: LOW, TO THE AMINO-TERMINAL DOMAIN OF GLUCOSE-6-  
CC PHOSPHATE ISOMERASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Y00459; CAA68518.1; -  
CC PIR: S17776; S17776.  
KW Streptomycin biosynthesis.  
SQ SEQUENCE 281 AA; 31726 MW; 75CEB24C CRC32;

Query Match 60.3%; Score 35; DB 1; Length 281;  
Best Local Similarity 70.0%; Pred. No. 6.5;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRIALR 11  
|:|:|:|:|  
Db 234 RLAAARLARM 243

## RESULT 3

DYHC\_NEUCR STANDARD; PRT; 4367 AA.  
ID DYHC\_NEUCR  
AC P45443;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN RO-1.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;

RX MEDLINE; 95014704.  
RA PLAMANN M., MINKE P.F., TINSLEY J.H., BRUNO K.S.;  
RT "Cytoplasmic dynein and actin-related protein Arp1 are required for  
RT normal nuclear distribution in filamentous fungi.";  
RL J. Cell Biol. 127:139-149(1994).  
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
CC NUCLEAR DISTRIBUTION IN HYPHAE.  
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
CC INTERMEDIATE AND LIGHT CHAINS.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L31504; AAA64908.1; -  
CC DR Motor protein; Microtubules; Dynein; ATP-binding;  
KW Heptad repeat pattern.  
FT NP\_BIND 1943 1950 ATP (POTENTIAL).  
FT NP\_BIND 2240 2247 ATP (POTENTIAL).  
FT NP\_BIND 2605 2612 ATP (POTENTIAL).  
FT NP\_BIND 2947 2954 ATP (POTENTIAL).  
FT SEQUENCE 4367 AA; 495568 MW; B81B5E92 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 4367;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 AIRRIALY 12  
|:|:|:|:|  
Db 2859 AVRIAMEY 2867

## RESULT 4

RL17\_CHLTR STANDARD; PRT; 122 AA.  
ID RL17\_CHLTR  
AC P47760;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 50S RIBOSOMAL PROTEIN L17.  
GN RPLQ.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=434/BU / SEROVAR L2;  
RX MEDLINE; 95247702.  
RA GU L.J., WENMAN W.M., REMACHA M., MEUSER R.U., COFFIN J.M., KAUL R.;  
RT "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and  
RT structural analysis.";  
RL J. Bacteriol. 177:2594-2601(1995).  
CC -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L33834; AAA74990.1; -  
CC DR PROSITE; PS01167; RIBOSOMAL\_L17; 1.  
DR PFAM; PF01196; Ribosomal\_L17; 1.

KW Ribosomal protein.  
SQ SEQUENCE 122 AA; 13969 MW; B8C43F7D CRC32;

Query Match 56.9%; Score 33; DB 1; Length 122;  
Best Local Similarity 54.5%; Pred. No. 6.5;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLAIIRIALRY 12  
|||:|:|:|  
Db 45 RLAVGLMWRY 55

RESULT 5  
TSMT\_MOUSE STANDARD; PRT; 264 AA.  
AC P40936;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE THIOETHER S-METHYLTRANSFERASE (EC 2.1.1.96) (TEMT).  
GN TEMT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RA WARNER D.R., HOFFMAN J.L.;  
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA WARNER D.R.;  
RL Thesis (1992), University of Louisville, U.S.A.  
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DIMETHYL SULFIDE -  
CC S-ADENOSYL-L-HOMOCYSTEINE + TRIMETHYLSULFONIUM.  
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS EC 2.1.1.1,  
CC EC 2.1.1.28 AND EC 2.1.1.96.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M88694; AAA62365.1; -;  
DR MGD; MGI:102963; TEMT.  
DR PROSITE; PS01100; NNMT\_PNMT\_TMT; 1.  
DR PFAM; PF01234; NNMT\_PNMT\_TMT; 1.  
KW Transferase; Methyltransferase.  
SQ SEQUENCE 264 AA; 29460 MW; 31B87F7A CRC32;

Query Match 56.9%; Score 33; DB 1; Length 264;  
Best Local Similarity 66.7%; Pred. No. 15;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLAIRRIA 9  
|||:|:|:|  
Db 177 YRAALRLA 185

RESULT 6  
BGL2\_BACSU STANDARD; PRT; 477 AA.  
AC P42403;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PROBABLE BETA-GLUCOSIDASE (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE)  
DE (BETA-D-GLUCOSIDE GLUCOHYDROLASE) (AMYGALASE).  
GN YCKE.

OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=168;  
RX MEDLINE; 95219080.  
RA FUJISHIMA Y., YAMANE K.;  
RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)  
RT of srfAA of the Bacillus subtilis chromosome.";  
RL Microbiology 141:277-279(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=168;  
RX MEDLINE; 97124189.  
RA YAMANE K., KUMANO M., KURITA K.;  
RT "The 25 degrees-36 degrees region of the Bacillus subtilis  
RT chromosome: determination of the sequence of a 146 kb segment and  
RT identification of 113 genes.";  
RL Microbiology 142:3047-3056(1996).  
RN [3]  
RP SEQUENCE OF 461-477 FROM N.A.  
RX MEDLINE; 88298684.  
RA VOSMAN B., KUIKEN G., KOOSTRA J., VENEMA G.;  
RT "Transformation in Bacillus subtilis: involvement of the  
RT 17-kilodalton DNA-entry nuclease and the competence-specific  
RT 18-kilodalton protein.";  
RL J. Bacteriol. 170:3703-3710(1988).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
CC GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D30762; BAA06429.1; -;  
DR EMBL; D50453; BAA08975.1; -;  
DR EMBL; M21672; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; Z99105; CAB2135.1; -;  
DR HSSP; P11546; 1PRG  
DR SUBTILIST; BG1181; YCKE.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_FL\_1; 1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_FL\_2; 1.  
DR PFAM; PF00232; Glyco\_hydro\_1; 1.  
KW Hypothetical protein; Hydrolase; Glycosidase; Cellulose degradation.  
FT ACT\_SITE 170 170 PROTON DONOR (POTENTIAL)  
FT ACT\_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 477 AA; 55140 MW; E92CD679 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 477;  
Best Local Similarity 54.5%; Pred. No. 29;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLAIIRIALRY 12  
|||:|:|:|  
Db 360 RIGLRITRY 370

RESULT 7  
PRC\_ECOLI STANDARD; PRT; 682 AA.  
ID PRC\_ECOLI  
AC P23855;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE TAIL-SPECIFIC PROTEASE PRECURSOR (EC 3.4.21.-) (PROTEASE RE) (PRC  
DE PROTEIN).

GN PRC OR TSP.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / W3110;  
 RX MEDLINE; 91310589.  
 RA HARA H., YAMAMOTO Y., HIGASHITANI A., SUZUKI H., NISHIMURA Y.;  
 RT "Cloning, mapping, and characterization of the Escherichia coli prc  
 RT gene, which is involved in C-terminal processing of penicillin-binding  
 RT protein 3";  
 RN J. Bacteriol. 173:4799-4813(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92108041.  
 RA SILBER K.R., KEILER K.C., SAUER R.T.;  
 RT "Tsp: a tail-specific protease that selectively degrades proteins  
 RT with nonpolar C termini";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:295-299(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE; 97426617.  
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 97251358.  
 RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T.,  
 RA ISONO K., KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M.,  
 RA MAKINO S., MIKI T., MIYOSHI K., MORI H., MORI T., MOTOMURA K.,  
 RA NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,  
 RA SAITO N., SANPEI G., SEKI Y., SIVASUNDARAM S., TAGAMI H.,  
 RA TAKEDA J., TAKEMOTO K., WADA C., YAMAMOTO Y., HORIUCHI T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40.1-50.0 min region on the linkage map";  
 RL DNA Res. 3:379-392(1996).  
 RN [5]  
 RP SEQUENCE OF 330-403 FROM N.A.  
 RX MEDLINE; 93077473.  
 RA SEANE A., SABBAG A., MCMURRY L.M., LEVY S.B.;  
 RT "Multiple antibiotic susceptibility associated with inactivation of  
 RT the prc gene";  
 RL J. Bacteriol. 174:7844-7847(1992).  
 RN [6]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 99175453.  
 RA KUNTE H.J., CRANE R.A., CULHAM D.E., RICHMOND D., WOOD J.M.;  
 RT "Protein ProQ influences osmotic activation of compatible solute  
 RT transporter Prop in Escherichia coli K-12";  
 RL J. Bacteriol. 181:1537-1543(1999).  
 RN [7]  
 RP ACTIVE SITE RESIDUES.  
 RX MEDLINE; 96081954.  
 RA KEILER K.C., SAUER R.T.;  
 RT "Identification of active site residues of the Tsp protease";  
 RL J. Biol. Chem. 270:28864-28868(1995).  
 RN [8]  
 RP SUBSTRATE SPECIFICITY.  
 RX MEDLINE; 96161995.  
 RA KEILER K.C., SAUER R.T.;  
 RT "Sequence determinants of C-terminal substrate recognition by the Tsp  
 RT protease";  
 RL J. Biol. Chem. 271:2589-2593(1996).  
 CC -1- FUNCTION: INVOLVED IN THE CLEAVAGE OF A C-TERMINAL PEPTIDE OF 11  
 CC RESIDUES FROM THE PRECURSOR FORM OF PENICILLIN-BINDING PROTEIN 3

(PBP3). PROTEASE THAT SELECTIVELY DEGRADES PROTEINS WITH NONPOLAR  
 C-TERMINAL. MAY BE INVOLVED IN PROTECTION OF THE BACTERIUM FROM  
 THERMAL AND OSMOTIC STRESSES.  
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE PERIPLASMIC SIDE OF THE  
 CC CYTOPLASMIC MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S41 (SERINE PROTEASE).  
 CC -1- SIMILARITY: CONTAINS 1 PDZ DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D00674; BAA00577.1;  
 DR EMBL; D00674; BAA00578.1; ALT\_INIT.  
 DR EMBL; M75634; AAA24699.1;  
 DR EMBL; AE000277; AAC74900.1;  
 DR EMBL; D90826; CAB21562.1;  
 DR EMBL; D90827; CAB21572.1;  
 DR EMBL; S49803; AAB24313.1;  
 DR EMBL; L48409; AAD41528.1;  
 DR PIR; A42475; A42475.  
 DR ECGENE; EG10760; PRC.  
 DR PFAM; PF00595; PDZ; 1.  
 DR Periplasmic; Hydrolase; Serine protease; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 682  
 FT DOMAIN 238 322  
 FT ACT\_SITE 452 452  
 FT ACT\_SITE 477 477  
 FT CONFLICT 317 317 L -> Q (IN REF. 1).  
 SQ SEQUENCE 682 AA; 76663 MW; 3CF7B39A CRC32;  
 Query Match 56.9%; Score 33; DB 1; Length 682;  
 Best Local Similarity 66.7%; Pred. No. 44;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 YRLAIRRIA 9  
 I: |||||  
 Db 195 YKFAIRRLA 203  
 RESULT 8  
 YPHG\_ECOLI ID YPHG\_ECOLI STANDARD; PRT; 1124 AA.  
 AC P76585;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 127.3 KD PROTEIN IN CSIE-GLYA INTERGENIC REGION.  
 GN YPHG.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE; 97426617.  
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AE000341; AAC75602.1; -  
 DR ECOGENE: EGI3468; yphg.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1124 AA; 127284 MW; 0A06B4C6 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 1124;  
 Best Local Similarity 87.5%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLAIRI 8  
 Db 412 YRLAIRI 419

RESULT 9  
 ID NH10\_YEAST STANDARD; PRT; 203 AA.  
 AC Q03435;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE NH10 PROTEIN.

GN NH10 OR HMO2 OR YDL002C OR YD8119.05C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomycetes.  
 [1]

RN SEQUENCE FROM N.A.

RP STRAIN=S288C / AB972;

RC MURPHY L., RICHARDS C., GENTLES S., HARRIS D., BARRELL B.G.,  
 RA RAJANDREAM M.A.;

RL Subitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

CC 1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC 1- SIMILARITY: CONTAINS 1 HMG BOX

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: Z48008; CAA88059.1; -

DR HSSP: P07155; LHME.

DR SGI: L0002765; NH10.

DR PFAM: PF00505; HMG\_box; 1.

KW Nuclear protein; DNA-binding. HMG BOX.

FT DNA\_BIND 94 158

SQ SEQUENCE 203 AA; 23857 MW; 80A216B6 CRC32;

Query Match 55.2%; Score 32; DB 1; Length 203;  
 Best Local Similarity 45.5%; Pred. No. 18;  
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLAIRIALRY 12  
 Db 28 RLKSVRLKLEY 38

RESULT 10

ID HRMA\_PSESY STANDARD; PRT; 375 AA.

AC Q08370;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE HRMA PROTEIN.  
 GN HRMA.

OS Pseudomonas syringae (pv. syringae).

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;

OC Pseudomonas.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PSS61;

RA HUANG H.C., HUTCHESON S.W., COLLIER A.;

RT "Characterization of the hrp cluster from Pseudomonas syringae pv.

RT syringae 61 and hrpA tagging of exported or membrane-spanning Hrp

RT proteins.";

RL Mol. Plant Microbe Interact. 4:469-476(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=PSS61;

RX MEDLINE; 94100578.

RA HEU S., HUTCHESON S.W.;

RT "Nucleotide sequence and properties of the hrma locus associated with

RT the Pseudomonas syringae pv. syringae 61 hrp gene cluster.";

RL Mol. Plant Microbe Interact. 6:553-564(1993).

CC 1- FUNCTION: UNKNOWN. MAY SERVE A REGULATORY FUNCTION.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: L14926; AA16545.1; -

KW Hypersensitive response.

SQ SEQUENCE 375 AA; 41458 MW; 733EB06 CRC32;

Query Match 55.2%; Score 32; DB 1; Length 375;  
 Best Local Similarity 58.3%; Pred. No. 36;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 YRLAIRIALRY 12

Db 174 YRLSIRTKTISY 185

RESULT 11

SECY\_BACSU

ID SECY\_BACSU STANDARD; PRT; 431 AA.

AC P16336;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE PREPROTEIN TRANSLOCASE SECY SUBUNIT.

GN SECY.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE; 90251170.

RA SUH J.-W., BOYLAN S.A., THOMAS S.M., DOLAN K.M., OLIVER D.B.,

RA PRICE C.W.;

RT "Isolation of a secY homologue from Bacillus subtilis: evidence for a

RT common protein export pathway in eubacteria.";

RL Mol. Microbiol. 4:305-314(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 90221911.

RA YOSHIKAWA H., DOI R.H.;

RT "Sequence of the Bacillus subtilis spectinomycin resistance gene

RT region.";

RL Nucleic Acids Res. 18:1647-1647(1990).

RN [3] SEQUENCE FROM N.A.  
RX MEDLINE: 90292990.  
RA NAKAMURA K., NAKAMURA A., TAKAMATSU H., YOSHIKAWA H., YAMANE K.;  
RT "Cloning and characterization of a Bacillus subtilis gene homologous  
to E. coli secY";  
RL J. Biochem. 107:603-607(1990).  
RN [4] SEQUENCE FROM N.A.  
RX MEDLINE: 96186897.  
RA SUH J.W., BOYLAN S.A., OH S.H., PRICE C.W.;  
RT "Genetic and transcriptional organization of the Bacillus subtilis  
sec-alpha region";  
RL Gene 169:17-23(1996).  
RN CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE  
CC TO ALLOW THE TRANSLLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,  
CC BY FORMING PART OF A CHANNEL.  
CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT  
CC COMPRISE THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X51329; CAA35712.1; -  
DR EMBL: M31102; AAB59118.1; -  
DR EMBL: D00619; BAA00495.1; ALT\_INIT.  
DR EMBL: L47971; AAB08819.1; -  
DR EMBL: Z99104; CAB11912.1; -  
DR PIR: S08629; BWBSSY.  
DR PIR: S12683; S12683.  
DR SUBTILIS: BG10445; SECY.  
DR PROSITE: PS00755; SECY\_1; 1.  
DR PROSITE: PS00756; SECY\_2; 1.  
DR PFAM: PF00344; secY; 1.  
KW Protein transport; Translocation; Transmembrane.  
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 18 39 POTENTIAL.  
FT DOMAIN 40 58 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 59 80 POTENTIAL.  
FT DOMAIN 81 114 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 115 132 POTENTIAL.  
FT DOMAIN 133 147 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 148 167 POTENTIAL.  
FT DOMAIN 168 173 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 174 192 POTENTIAL.  
FT DOMAIN 193 216 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 217 234 POTENTIAL.  
FT DOMAIN 235 267 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 268 291 POTENTIAL.  
FT DOMAIN 292 309 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 310 329 POTENTIAL.  
FT DOMAIN 330 366 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 367 386 POTENTIAL.  
FT DOMAIN 387 391 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 392 410 POTENTIAL.  
FT DOMAIN 411 431 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 431 AA; 47243 MW; 1C8A4316 CRC32;

Query Match 55.2%; Score 32; DB 1; Length 431;  
Best Local Similarity 55.6%; Pred. No. 42;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 4 AIRIALRY 12  
|:|:|:|:  
DB 237 AVRKIAIQY 245

RESULT 12  
ACH6\_CAEEL STANDARD; PRT; 487 AA.  
ID ACH6\_CAEEL  
AC Q93149;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ACETYLCHOLINE RECEPTOR, BETA-TYPE SUBUNIT ACR-3 PRECURSOR.  
GN ACR-3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 98269564.  
RA BAYLIS H.A., MATSUDA K., SQUIRE M.D., FLEMING J.T., HARVEY R.J.,  
RA DARLISON M.G., BARNARD E.A., SATTELE D.B.;  
RT "ACR-3, a Caenorhabditis elegans nicotinic acetylcholine receptor  
subunit. Molecular cloning and functional expression";  
RL Recept. Channels 5:149-158(1997).  
CC -1- FUNCTION: POSSIBLE ACETYLCHOLINE RECEPTOR.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Y08637; CAA69927.1; -  
DR PROSITE: PS00236; NEUROTR\_ION\_CHANNEL; 1.  
DR PFAM: PF00065; neur\_chan; 1.  
KW Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein;  
KW Transmembrane; Signal;  
FT SIGNAL 1 487 POTENTIAL.  
FT CHAIN ? 487 ACETYLCHOLINE RECEPTOR, BETA-TYPE SUBUNIT  
FT ACR-3.  
FT DOMAIN ? 231 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 259 279 POTENTIAL.  
FT TRANSMEM 294 314 POTENTIAL.  
FT DOMAIN 315 439 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 440 460 POTENTIAL.  
FT CARBOHYD 46 46 POTENTIAL.  
FT DISULFID 151 165 BY SIMILARITY.  
SQ SEQUENCE 487 AA; 56245 MW; 102629B3 CRC32;

Query Match 55.2%; Score 32; DB 1; Length 487;  
Best Local Similarity 58.3%; Pred. No. 48;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLAIRRIARY 12  
|:|:|:|:|:  
DB 219 YQIKIRKALFY 230

RESULT 13  
PRXC\_CURIN STANDARD; PRT; 609 AA.  
ID PRXC\_CURIN  
AC P49053;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE VANADIUM CHLOROPEROXIDASE (EC 1.11.1.10) (VCPO) (VANADIUM CHLORIDE  
PEROXIDASE).  
GN CPO.

OS Curvularia inaequalis.  
 CC Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;  
 CC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Curvularia.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE; 95262722.  
 RA SIMONS B.H., BARNETT P., VOLLENBROEK E.G.M., DEKKER H.L.,  
 RA MULJERS A.O., MESSERSCHMIDT A., WEVER R.;  
 RT "Primary structure and characterization of the vanadium  
 RT chloroperoxidase from the fungus Curvularia inaequalis";  
 RL Eur. J. Biochem. 229:566-574(1995).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
 RX MEDLINE; 96133943.  
 RA MESSERSCHMIDT A., WEVER R.;  
 RT "X-ray structure of a vanadium-containing enzyme: chloroperoxidase  
 RT from the fungus Curvularia inaequalis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:392-396(1996).  
 CC -1- CATALYTIC ACTIVITY: 2 RH + 2 CHLORIDE + H(2)O(2) -> 2 RCL + 2  
 CC H(2)O.  
 CC -1- COFACTOR: VANADIUM.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X85369; CAA59686.1; -  
 DR PDB; 1VNC; 08-NOV-96. -  
 KW Oxidoreductase; Peroxidase; Vanadium; 3D-structure.  
 FT ACT\_SITE 404 404  
 FT METAL 496 496 VANADIUM.  
 FT CONFLICT 454 454 P -> S (IN AA SEQUENCE).  
 SQ SEQUENCE 609 AA; 67530 MW; 19112E80 CRC32;  
 CC  
 Query Match 55.28; Score 32; DB 1; Length 609;  
 Best Local Similarity 50.08; Pred. No. 61;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 YRLAIRRIALRY 12  
 DB 304 YNQIVRRIAVTY 315  
 ID YE7A\_SCHPO STANDARD; PRT; 723 AA.  
 AC O14162;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 80.1 KD PROTEIN C4A8.10 IN CHROMOSOME 1.  
 GN SPAC4A8.10.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RA SKELTON J., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO YEAST YGL144C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; Z98762; CAB11480.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 723 AA; 80090 MW; 159D795B CRC32;  
 CC  
 Query Match 55.28; Score 32; DB 1; Length 723;  
 Best Local Similarity 54.58; Pred. No. 74;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 YRLAIRRIALR 11  
 DB 647 HRLAWRKVAVR 657  
 ID MDR3\_CRIGR STANDARD; PRT; 1281 AA.  
 AC P23174;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 01-NOV-1991 (Rel. 35, Last annotation update)  
 DE MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPROTEIN 3).  
 GN PGP3 OR PGP3.  
 OS Cricetus griseus (Chinese hamster).  
 CC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92135896.  
 RA ENDICOTT J.A., SARANGI F., LING V.;  
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein  
 RT gene family";  
 RL DNA Seq. 2:89-101(1991).  
 CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED  
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT  
 CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY  
 CC CANNOT.  
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 CC (ABC TRANSPORTERS). MDR SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M60042; AAA68885.1; -  
 DR HSSP; P13569; INED  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR PFAM; PF00005; ABC\_tran; 2.  
 DR PFAM; PF00664; ABC\_membrane; 2.  
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Duplication;  
 KW Multigene family.  
 FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 58 78 POTENTIAL.  
 FT TRANSMEM 122 142 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 218 238 POTENTIAL.  
 FT TRANSMEM 299 319 POTENTIAL.  
 FT TRANSMEM 328 348 POTENTIAL.  
 FT TRANSMEM 349 712 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 713 733 POTENTIAL.  
 FT TRANSMEM 758 778 POTENTIAL.  
 FT TRANSMEM 834 854 POTENTIAL.

FT TRANSMEM 855 875 POTENTIAL.  
FT TRANSMEM 938 958 POTENTIAL.  
FT TRANSMEM 975 995 POTENTIAL.  
FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).  
FT NP\_BIND 429 436 ATP (POTENTIAL).  
FT NP\_BIND 1071 1078 ATP (POTENTIAL).  
SQ SEQUENCE 1281 AA; 140866 MW; 9FE9DF5C CRC32;

Query Match 55.2%; Score 32; DB 1; Length 1281;  
Best Local Similarity 70.0%; Pred. NO. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLAIIRIALR 11  
Db 1184 RLAIIRIALR 1193

Search completed: February 8, 2000, 01:26:02  
Job time: 1562 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 19:16:17 ; Search time 176.54 Seconds

(without alignments)  
4.713 Million cell updates/sec

Title: US-08-653-294-36

Perfect score: 58

Sequence: 1 YRLAIRRIALRY 12

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL\_12:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	70.7	795	5 Q22063	Q22063 caenorhabdi
2	40	69.0	309	2 O86347	O86347 mycobacteri
3	37	63.8	243	2 O85853	O85853 sphingomona
4	37	63.8	325	2 O9XB05	O9XB05 myxococcus
5	36	62.1	313	2 P97213	P97213 clostridium
6	36	62.1	441	5 Q26954	Q26954 trypanosoma
7	35	60.3	276	2 P72880	P72880 synecocyst
8	35	60.3	281	2 Q54259	Q54259 streptomyce
9	35	60.3	436	2 O9X876	O9X876 streptomyce
10	35	60.3	725	5 O16382	O16382 caenorhabdi
11	34	58.6	178	1 Q9YAA5	Q9YAA5 aeropyrum p
12	34	58.6	267	2 Q53815	Q53815 streptomyce
13	34	58.6	267	2 Q54257	Q54257 streptomyce
14	34	58.6	350	5 O45803	O45803 caenorhabdi
15	34	58.6	384	10 Q9ZUJ4	Q9ZUJ4 arabidopsis
16	34	58.6	469	1 O28589	O28589 archaeoglob
17	34	58.6	573	1 Q9YDG3	Q9YDG3 aeropyrum p
18	34	58.6	1899	10 Q9XEG1	Q9XEG1 gossypium h
19	33	56.9	128	3 P87286	P87286 saccharomyc
20	33	56.9	141	2 O84514	O84514 chlamydia t

21	33	56.9	190	2 O67313	O67313 aquifex aeo
22	33	56.9	266	2 Q9Z4B6	Q9Z4B6 escherichia
23	33	56.9	285	5 P91088	P91088 caenorhabdi
24	33	56.9	321	5 O45795	O45795 caenorhabdi
25	33	56.9	404	10 Q9XIL2	Q9XIL2 arabidopsis
26	33	56.9	452	2 Q59958	Q59958 streptococc
27	33	56.9	518	2 Q49843	Q49843 mycobacteri
28	33	56.9	530	12 Q9YVN7	Q9YVN7 melanoplus
29	33	56.9	904	4 O15455	O15455 homo sapien
30	32	55.2	61	12 Q9WHV4	Q9WHV4 squash yell
31	32	55.2	132	2 Q9X378	Q9X378 bacillus an
32	32	55.2	142	2 Q9Z7S9	Q9Z7S9 chlamydia p
33	32	55.2	215	5 O21162	O21162 caenorhabdi
34	32	55.2	259	11 Q60503	Q60503 cricetus
35	32	55.2	269	5 O16763	O16763 caenorhabdi
36	32	55.2	306	2 Q9WZ99	Q9WZ99 thermotoga
37	32	55.2	339	1 Q9YA89	Q9YA89 aeropyrum p
38	32	55.2	399	5 Q26952	Q26952 trypanosoma
39	32	55.2	407	10 Q23555	Q23555 arabidopsis
40	32	55.2	434	10 Q49569	Q49569 arabidopsis
41	32	55.2	442	2 P95486	P95486 pseudomonas
42	32	55.2	490	2 O07074	O07074 porphyromon
43	32	55.2	573	5 Q97211	Q97211 leishmania
44	32	55.2	828	10 O82630	O82630 arabidopsis
45	32	55.2	978	4 O60313	O60313 homo sapien

#### ALIGNMENTS

RESULT 1

ID Q22063 PRELIMINARY; PRT: 795 AA.  
AC Q22063; Q93489;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE T01C3.10 PROTEIN.  
GN T01C3.10.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]  
RP SEQUENCE FROM N.A.  
RA WILD A.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z81061; CAB02939.1; -

DR EMBL; Z78413; CAB02939.1; JOINED.

DR EMBL; Z78413; CAB01667.1; -

DR EMBL; Z81061; CAB01667.1; JOINED.

DR HSSP; P19491; IGR2

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

DR PFAM; PF00060; lig\_chan; 1.

OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 ON Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;

RX MEDLINE; 98295987.  
 RA COLE S.T., BROSH R., PARKHILL J., GARNIER T., CHURCH C., HARRIS D.,  
 RA GORDON S.V., EIGLMER K., GAS S., BARRY III C.E., TEKAIA F.,  
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,  
 RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAWLIN N., HOLROYD S.,  
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,  
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,  
 RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SOARES R., SULSTON J.E.,  
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.

RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."  
 RL Nature 393:537-544(1998).

RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN-H37RV;  
 RA PARKHILL J.

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL008967; CAA15591.1; -  
 DR HSSP; P33164; 2PIA.

DR PROSITE; PS00197; 2FE2S-FERREDOXIN; 1.

DR PFAM; PF00111; fer2; 1.

DR PFAM; PF00175; oxidored\_fad; 1.

KW Hypothetical protein; Iron-sulfur.

SQ SEQUENCE 309 AA; 33517 MW; B152B590 CRC32;

Query Match 69.0%; Score 40; DB 2; Length 309;  
 Best Local Similarity 88.9%; Pred. No. 4.5;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLAIRRIA 9  
 ||| |||||  
 DB 65 YRLAIRRIA 73

RESULT 3

ID O85853 PRELIMINARY; PRT; 243 AA.

AC O85853;

DT 01-NOV-1998 (T-EMBLrel. 08, Created)

DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)

DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)

DE HYPOTHETICAL 26.5 KD PROTEIN.

OS Sphingomonas aromaticivorans.

OG Plasmid pNL1.

OC Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;

OC Sphingomonas.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-F199;

RA ROMINE M.F., STILLWELL L.C., WONG K.-K., THURSTON S.J., SISK E.C.,

RA SENSEN C.W., GAASTERLAND T., SAFFER J.D., FREDRICKSON J.K.;

RT "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas

aromaticivorans strain F199."

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF079317; AAD03868.1; -

KW Hypothetical protein; Plasmid.

SQ SEQUENCE 243 AA; 26455 MW; 40CDFB4 CRC32;

Query Match 63.8%; Score 37; DB 2; Length 243;  
 Best Local Similarity 63.6%; Pred. No. 13;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLAIRRIALRY 12

||| |||: ||

DB 5 RLQRRVTIRY 15

RESULT 4

Q9XB05 PRELIMINARY; PRT; 325 AA.

ID Q9XB05;

DT 01-NOV-1999 (T-EMBLrel. 12, Created)

DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)

DE MEMBRANE ASSOCIATED PROTEIN.

GN TAD.

OS Myxococcus xanthus.

OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;

OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ER-15;

RA PAITAN Y., ORR E., RON E.Z., ROSENBERG E.;

RT "Genetic and functional analysis of genes required for the post-

RT modification of the polyketide antibiotic TA of Myxococcus xanthus."

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ132503; CAB46503.1; -

SQ SEQUENCE 325 AA; 35985 MW; 4CC64E85 CRC32;

Query Match 63.8%; Score 37; DB 2; Length 325;

Best Local Similarity 58.3%; Pred. No. 18;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12

||| : |||

DB 202 YRLTVDRFPLRY 213

RESULT 5

P97213

ID P97213 PRELIMINARY; PRT; 313 AA.

AC P97213;

DT 01-MAY-1997 (T-EMBLrel. 03, Created)

DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)

DT 01-MAY-1997 (T-EMBLrel. 03, Last annotation update)

DE CDU2, CDU1, TCDD, TCDB, TCDE, TCDA, TCDC, CDD1, CDD2, CDD3, AND CDD4

DE GENES.

GN CDU2.

OS Clostridium difficile.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VP110463;

RA VON EICHEL-STREIBER C.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; X92982; CAA63559.1; -

DR EMBL; X92982; CAA63558.1; -

SQ SEQUENCE 313 AA; 33380 MW; 4F20347A CRC32;

Query Match 62.1%; Score 36; DB 2; Length 313;

Best Local Similarity 41.7%; Pred. No. 27;

Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLAIRRIALRY 12

||| : |||

DB 259 YRVVARLSIKY 270

RESULT 6

Q26954

ID Q26954 PRELIMINARY; PRT; 441 AA.

AC Q26954;

DT 01-NOV-1996 (T-EMBLrel. 01, Created)

DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)

DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)

DE TCJ4.  
 GN Trypanosoma cruzi.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PBOL;  
 RX MEDLINE: 98226175.  
 RA TIBBETTS R.S., JENSEN J.L., OLSON C.L., WANG F., ENGMAN D.M.;  
 RT "The DnaJ family of protein chaperones in Trypanosoma cruzi."  
 RL EMBL: L46819; AAC18897.1; -  
 DR HSSP: P08622; IXL.  
 DR PROSITE: PS00636; DNAJ\_1.1.  
 DR PFAM: PF00226; DnaJ\_1.  
 SQ SEQUENCE 441 AA; 48819 MW; E6AD2FA2 CRC32;

Query Match 62.1%; Score 36; DB 5; Length 441;  
 Best Local Similarity 63.6%; Pred. No. 38;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 RLAIIRIALRY 12  
 | | | | | | | | | |  
 Db 23 RTAYRLALRY 33

RESULT 7  
 P72880 PRELIMINARY; PRT; 276 AA.  
 AC P72880;  
 DT 01-FEB-1997 (T-EMBLrel. 02, Created)  
 DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)  
 DE 01-JAN-1999 (T-EMBLrel. 09, Last annotation update)  
 DE ABC TRANSPORTER.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC6803;  
 RA TABATA S.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC6803;  
 RX MEDLINE: 97061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., NAKAZAKI N., NARUO K., OKUMURA S.,  
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
 RA TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 entire genome and assignment of potential protein-coding regions."  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL: D90901; BAA16896.1; -  
 DR PFAM: PF01061; ABC2\_membrane; 1.  
 SQ SEQUENCE 276 AA; 31542 MW; 4DBF5330 CRC32;

Query Match 60.3%; Score 35; DB 2; Length 276;  
 Best Local Similarity 66.7%; Pred. No. 37;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 YRLAIRIALRY 12  
 | | | | | | | | | |  
 Db 31 YTLAWRDIAVRY 42

RESULT 8  
 Q54259 PRELIMINARY; PRT; 281 AA.  
 ID Q54259  
 AC Q54259;

DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)  
 DE GLA 0 GENES STRB1, STRF, STRG, STRH, STRP.  
 GN STRF.  
 OS Streptomyces glaucescens.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GLA 0;  
 RA MAYER G., PIEPERSBERG W.;  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 (GLA 0);  
 RA PIEPERSBERG W.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 (GLA 0);  
 RA RETZLAFF L., MAYER G., BEYER S., AHLERT J., VERSECK S., DISTLER J.,  
 RA PIEPERSBERG W.;  
 RT "Streptomycin Production in Streptomyces: a Progress Report."  
 RL (In) Baltz R.H., Hegeman G.D., Skatrud P.L. (eds.);  
 RL Industrial microorganisms. Basic and applied molecular genetics,  
 RL pp.183-194, ASM Press, Herndon (1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 (GLA 0);  
 RX MEDLINE: 96204519.  
 RA BEYER S., DISTLER J., PIEPERSBERG W.;  
 RT "The str gene cluster for the biosynthesis of 5'-hydroxystreptomycin  
 in Streptomyces glaucescens GLA.0 (ETH 22794): new operons and  
 evidence for pathway-specific regulation by StrR."  
 RL Mol. Gen. Genet. 250:775-784(1996).  
 DR EMBL: X78974; CAA55572.1; -  
 DR EMBL: AJ006985; CAA07379.1; -  
 SQ SEQUENCE 281 AA; 31427 MW; 30CD0C63 CRC32;

Query Match 60.3%; Score 35; DB 2; Length 281;  
 Best Local Similarity 70.0%; Pred. No. 38;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 RLAIIRIALR 11  
 | | | | | | | | | |  
 Db 234 RIAARRLALR 243

RESULT 9  
 Q9X876 PRELIMINARY; PRT; 436 AA.  
 ID Q9X876  
 AC Q9X876;  
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN (FRAGMENT).  
 GN SCE15.01.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MURPHY L., HARRIS D.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 DR BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE; 97000351.  
RA KINASHI H., HOPWOOD D.A.;  
RA REDENBACH M., KIESER H.M., DENAPATTE D., EICHNER A., CULLUM J.,  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL MOL. Microbiol. 21:77-96(1996).  
DR EMBL; AL049707; CAB41270.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 436 436  
SQ SEQUENCE 436 AA; 46734 MW; EDFD0658 CRC32;

Query Match 60.3%; Score 35; DB 2; Length 436;  
Best Local Similarity 70.0%; Pred. No. 59;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 RLAIIRIALR 11  
Db 123 RLAVRTALR 132  
|||:|:|:|

RESULT 10

OL6382 PRELIMINARY; PRT; 725 AA.  
AC OL6382;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DE K12B6.2 PROTEIN.  
DE K12B6.2.  
GN K12B6.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA JONES K., WOLDMANN P.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016442; AAB65917.1; -.  
SQ SEQUENCE 725 AA; 84926 MW; A4847D75 CRC32;

Query Match 60.3%; Score 35; DB 5; Length 725;  
Best Local Similarity 58.3%; Pred. No. 99;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 YRLAIRIALRY 12

Db 594 FSFAIRIVLVNY 605  
: ||||| |

RESULT 11

ID Q9YAA5 PRELIMINARY; PRT; 178 AA.  
AC Q9YAA5;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DE 178AA LONG HYPOTHETICAL PROTEIN.  
GN APE2034.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE; 99310339.  
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOVAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
crenarchaeon, aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000063; BAA81044.1; -.  
SQ SEQUENCE 178 AA; 19484 MW; FEA2CD00 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 178;  
Best Local Similarity 63.8%; Pred. No. 37;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 2 RLAIIRIALRY 12  
Db 34 RMACRIATVY 44  
||:||||:|

RESULT 12

Q53815 PRELIMINARY; PRT; 267 AA.  
ID Q53815;  
AC Q53815;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE ISP 5564 GENES STR AND STRF (FRAGMENT).  
GN STRF.  
OS Streptomyces blausensis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomyceae; Streptomycetaceae; Streptomyces.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISP 5564;  
RA MAYER G., PIEPERSBERG W.;  
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X78972; CAA55568.1; -.  
FT NON\_TER 267 267  
SQ SEQUENCE 267 AA; 29777 MW; CBC0D606 CRC32;

Query Match 58.6%; Score 34; DB 2; Length 267;  
Best Local Similarity 70.0%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 RLAIIRIALR 11  
Db 234 RVAARLALR 243  
||:||||

RESULT 13

Q54257  
ID Q54257 PRELIMINARY; PRT; 267 AA.  
AC Q54257;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMELrel. 01, Last annotation update)  
DE DSM 40480 GENES STRB1 AND STRF (FRAGMENT).  
GN STRF.  
OS Streptomyces galbus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DSM 40480;  
RA MAYER G., PIEPERSBERG W.;  
RL Submitted (APR-1994) to the EMBL/GenBank/DBSJ databases.  
DR EMBL: X78973; CAA55570.1;  
FT NON\_TER 267  
SQ SEQUENCE 267 AA; 29909 MW; 82D9D7A4 CRC32;  
  
Query Match 58.6%; Score 34; DB 2; Length 267;  
Best Local Similarity 70.0%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 RLAIIRIALR 11  
||| ||:|:|  
DB 234 RLAAARLAVR 243  
  
RESULT 14  
O45803 PRELIMINARY; PRT; 350 AA.  
AC O45803;  
DT 01-JUN-1998 (TREMELrel. 06, Created)  
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)  
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)  
DE T23D5.1 PROTEIN.  
GN T23D5.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA LLOYD C., WILKINSON J.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBSJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL: 282051; CAB04815.1;  
DR PFAW: PF01461; 7tm\_4; 1.  
SQ SEQUENCE 350 AA; 40291 MW; A9117B4D CRC32;

Query Match 58.6%; Score 34; DB 5; Length 350;  
Best Local Similarity 75.0%; Pred. No. 73;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YRLAIRRI 8  
||||:|:

DB 307 YRLAVRKI 314  
  
RESULT 15  
Q92UJ4  
ID Q92UJ4 PRELIMINARY; PRT; 384 AA.  
AC Q92UJ4;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE T2K10.3 PROTEIN.  
GN T2K10.3.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA VYSOTSKAIA V.S., SCHWARTZ J.R., YU G., TORIUMI M., LENZ C., LIU S.,  
RA LI J., KREMENTSKAIA I., LUIROS J., GONZALEZ A., ALTAFI H., ARAUJO R.,  
RA RUEHLER E., CHAO Q., CONN L., CONWAY A.B., DUNN P., HANSEN N.,  
RA HUIZAR L., KIM C., PALM C., ROWLEY D., SHINN P., WALKER M.,  
RA DAVIS R.W., ECKER J.R., FEDERSPIEL N.A., THEOLOGIS A.;  
RT "Arabidopsis thaliana chromosome 1 BAC T2K10 sequence.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBSJ databases.  
DR EMBL: AC005966; AAD14474.1;  
DR HSSP: P25685; 1HDJ.  
DR PROSITE: PS00636; DNAJ\_1; 1.  
SQ SEQUENCE 384 AA; 43306 MW; BB0D711E CRC32;  
  
Query Match 58.6%; Score 34; DB 10; Length 384;  
Best Local Similarity 77.8%; Pred. No. 81;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 AIRRIALRY 12  
|:|:|:|  
DB 42 AYRERVALRY 50  
  
Search completed: February 8, 2000, 19:16:19  
Job time: 21506 sec

**THIS PAGE BLANK (USPTO)**



Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 8849)

Adams,M. and Venter,J.C.

Direct Submission

Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA

This sequence was identified as CDM:10212817 by the submitter.  
For more information on this record e-mail to flycelera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

source

1..8849

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 2604 a 1735 c 1894 g 2626 t

ORIGIN

alignment\_scores:

Quality: 46.00 Length: 12  
Ratio: 3.833 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 66.667

alignment\_block:

US-08-653-294-36 x AC017966

Align seg 1/1 to: AC017966 from: 1 to: 8849

1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12

844 TACAGAATGCGATGCAACGATGCTTACGTTAT 879

seq\_name: gb\_h94:AC010564

seq\_documentation\_block:

LOCUS AC010564 143914 bp DNA HTG 16-OCT-1999  
DEFINITION Drosophila melanogaster chromosome 3L/62A1 clone RPC198-2701, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 89 unordered pieces.

ACCESSION AC010564

VERSION AC010564.4 GI:5917942

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 143914)

Murny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,

Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhray,C., Bunac,C.,

Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,

David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,

Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,

Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,

Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hognes,M.,

Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,

Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,

Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,

Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,

Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,

Parish,B., Paxton,S., Payton,B., Perez,L., Pu.L.L., Quiles,M.,

Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shan,E.,

Shen,H., Simon,M., Sparks,A., Stamps,A., Sugchang,R., Tabor,P.,

Taylor,T., Vasquez,L., Vinson,R., Vo.Q., Wabba,M., Watlington,S.,

Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,

Wrenford,G., Yu.W., Zhou,X., Nelson,D. and Gibbs,R.

Direct Submission

Unpublished

REFERENCE 2 (bases 1 to 143914)

Worley,K.C.

TITLE  
JOURNAL

COMMENT

Direct Submission  
Submitted (16-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Sep 22, 1999 this sequence version replaced gi:5916428.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 89 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 831: contig of 831 bp in length

832: contig of 814 bp in length

1646: contig of 874 bp in length

2519: contig of 831 bp in length

3350: contig of 856 bp in length

4206: contig of 1263 bp in length

5469: contig of 844 bp in length

6313: contig of 960 bp in length

5470: contig of 828 bp in length

6314: contig of 840 bp in length

7274: contig of 828 bp in length

8102: contig of 840 bp in length

8942: contig of 875 bp in length

10145: contig of 875 bp in length

11020: contig of 806 bp in length

11825: contig of 1260 bp in length

13085: contig of 852 bp in length

13938: contig of 815 bp in length

14752: contig of 1220 bp in length

15972: contig of 599 bp in length

16571: contig of 613 bp in length

17184: contig of 834 bp in length

18018: contig of 819 bp in length

18837: contig of 1554 bp in length

20391: contig of 861 bp in length

21252: contig of 718 bp in length

21970: contig of 1056 bp in length

23026: contig of 1515 bp in length

24541: contig of 1185 bp in length

25726: contig of 844 bp in length

26571: contig of 1111 bp in length

27681: contig of 1549 bp in length

29230: contig of 1405 bp in length

30635: contig of 1122 bp in length

31757: contig of 1142 bp in length

32899: contig of 1193 bp in length

34092: contig of 836 bp in length

34928: contig of 777 bp in length

35705: contig of 1186 bp in length

36891: contig of 1512 bp in length

38403: contig of 1535 bp in length

39938: contig of 1219 bp in length

41157: contig of 848 bp in length

42005: contig of 1425 bp in length

43465: contig of 1036 bp in length

44890: contig of 1572 bp in length

45926: contig of 1602 bp in length

47498: contig of 1271 bp in length

49100: contig of 1605 bp in length

50372: contig of 2074 bp in length

51976: contig of 1977 bp in length

54050: contig of 1384 bp in length

56027: contig of 1317 bp in length

57411: contig of 1228 bp in length

58813: contig of 2134 bp in length

60132: contig of 1140 bp in length

61360: contig of 1221 bp in length

63494: contig of 979 bp in length

64634: contig of 1211 bp in length

65855: contig of 828 bp in length

66834: contig of 828 bp in length

68045: contig of 828 bp in length

68873: contig of 828 bp in length

68046



```

* 68874 70019: contig of 1146 bp in length
* 70020 72323: contig of 2304 bp in length
* 72324 73759: contig of 1436 bp in length
* 73760 75834: contig of 2075 bp in length
* 75835 78389: contig of 2555 bp in length
* 78390 80840: contig of 2451 bp in length
* 80841 82111: contig of 1271 bp in length
* 82112 84052: contig of 1941 bp in length
* 84053 86110: contig of 2058 bp in length
* 86111 87488: contig of 1378 bp in length
* 87489 89108: contig of 1620 bp in length
* 89109 91457: contig of 2349 bp in length
* 91458 92907: contig of 1450 bp in length
* 92908 95337: contig of 2430 bp in length
* 95338 97095: contig of 1758 bp in length
* 97096 100091: contig of 2996 bp in length
* 100092 102072: contig of 1981 bp in length
* 102073 104497: contig of 2425 bp in length
* 104498 107509: contig of 3011 bp in length
* 107510 109599: contig of 2090 bp in length
* 109600 112405: contig of 2807 bp in length
* 112406 115205: contig of 2799 bp in length
* 115206 118318: contig of 3113 bp in length
* 118319 121885: contig of 3568 bp in length
* 121886 125221: contig of 3336 bp in length
* 125222 129338: contig of 4117 bp in length
* 129339 132811: contig of 3473 bp in length
* 132812 137892: contig of 5081 bp in length
* 137893 143914: contig of 6022 bp in length.

```

```

FEATURES
  source
    1. 143914
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /chromosome="3L/62A1"
      /clone="RPC198-2701"
BASE COUNT 38720 a 33523 c 32072 g 39042 t 557 others
ORIGIN

```

```

alignment_scores
  Quality: 46.00 Length: 12
  Ratio: 3.833 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 66.667

```

```

alignment_block:
US-08-653-294-36 x AC010564

```

```

Align seg 1/1 to: AC010564 from: 1 to: 143914

```

```

1 TvrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||:.....:|||||:|||||:|||||:|||||:
58362 TACAGAATGTCGATGCACGCGATGCTTACGTTAT 58397

```

```

seq_name: gb_in2.AC005847

```

```

seq_documentation_block:
LOCUS AC005847 268369 bp DNA INV 22-OCT-1998
DEFINITION Drosophila melanogaster DNA sequence (Pls DS031179 (D226), DS06357 (D230), DS03404 (D231), DS06962 (D232), and DS07291 (D240)), complete sequence.
ACCESSION AC005847 AC004331 AC004436 AC004438 AC004489
VERSION AC005847.1 GI:3779012
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in tet from p1 clones DS03179 (D226), DS06357 (D230), DS03404 (D231), DS06962 (D232), and DS07291 (D240)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 268369)
AUTHORS Celisner, S.E., George, R.A., Galle, R.F., Hoskins, R.A., Swirskas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E.,

```

```

Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., PUNCH, E., Snir, E., Twomey, B., Wan, K.H., Zhang, R., Zieran, L.L. and Rubin, G.M. Sequencing of Drosophila chromosome 3L, region 61F3-62A2 Unpublished (1997)
2 (bases 1 to 268369)

```

# AUTHORS

```

Celisner, S.E., George, R.A., Galle, R.F., Hoskins, R.A., Swirskas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., PUNCH, E., Snir, E., Twomey, B., Wan, K.H., Zhang, R., Zieran, L.L. and Rubin, G.M. Direct Submission
Submitted (22-OCT-1998) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US

```

# COMMENT

```

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdg@fruitfly.berkeley.edu.
Library locations: 11-34, 21-67, 44-36, 73-50, 92-76.

```

# FEATURES

## source

```

1. 268369
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /chromosome="3L"
  /map="61F3-62A2"
  /clone="Pls DS031179 (D226), DS06357 (D230), DS03404 (D231), DS06962 (D232), and DS07291 (D240)"
  /note="These five plis were completed as a project. The coordinates are the pl ends for each pl. DS03404 (D231) extends from bp 1 to bp 80,732. DS031179 (D226) extends from bp 32,734 to bp 117,954. DS06357 (D230) extends from bp 86,745 to bp 169,482. DS06962 (D232) extends from bp 127,117 to bp 209,427. DS07291 (D240) extends from bp 194,759 to bp 268,369."
BASE COUNT 71701 a 61029 c 61736 g 73903 t
ORIGIN

```

## alignment\_scores:

```

  Quality: 46.00 Length: 12
  Ratio: 3.833 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 66.667

```

## alignment\_block:

```

US-08-653-294-36 x AC005847/rev

```

```

Align seg 1/1 to reverse of: AC005847 from: 1 to: 268369

```

```

1 TvrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||:.....:|||||:|||||:|||||:|||||:
46162 TACAGAATGTCGATGCACGCGATGCTTACGTTAT 46127

```

```

seq_name: gb_bai:MLCB2533

```

## seq\_documentation\_block:

```

LOCUS MLCB2533 40245 bp DNA BCT 27-AUG-1999
DEFINITION Mycobacterium leprae cosmid B2533.
ACCESSION AL035310
VERSION AL035310.1 GI:4200258
KEYWORDS ansp. ATP-dependent RNA helicase; ATP-phosphoribosyl transferase; ATPase; hnsG 5-methyltetrahydrofolate-homocysteine methyl transferase; hnsI; L-asparagine permease; meth; mtbB; phosphoribosyl-AMP cyclohydrolase; prcA; prcB; proteasome alpha subunit; proteasome beta subunit; protein translocation system;

```

pseudogene; RLEP; sec-independent.  
Mycobacterium leprae.  
Mycobacterium leprae  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium.  
1 (bases 1 to 40245)  
Eiglmeler K., Honore N., Woods S.A., Caudron B. and Cole S.T.  
Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae  
Mol. Microbiol. 7 (2), 197-206 (1993)  
93188700  
2 (bases 1 to 40245)  
Hamlin N. and Churcher C.M.  
Unpublished  
3 (bases 1 to 40245)  
James K.D., Parkhill J., Barrell B.G. and Rasthriam M.A.  
Direct Submission  
Submitted (26-JAN-1998) Mycobacterium leprae sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.  
Stewart T. Cole. [3] Unite de Genetique Moleculaire Bacterienne,  
Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,  
France Requests for cosmids should be sent to Karin Eiglmeler  
(kei@pasteur.fr)  
Notes:  
The Sanger Centre is funded to complete the sequence of M. leprae  
by the Heiser Program for Research in Leprosy and Tuberculosis of  
The New York Community Trust.  
Work in Paris is supported by the Heiser Trust, the Association  
Francaise Raoul Follereau and the Groupement de Recherches et des  
Etudes des Genomes (GIP-GREG).  
Details of M. leprae sequencing at the Sanger Centre are available  
on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/projects/>)  
CDS are numbered using the following system eg MLCB33.01c. ML (M.  
leprae), CB33 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon. All CDS  
over 100 codons have been analysed. Gene prediction is based on  
positional base preference in codons especially where there is an  
increase in the observed/expected third position G + C. CAUTION:  
We may not have predicted the correct initiation codon. Where  
possible we choose an initiation codon (atg, gtg, or ttg) which is  
preceded by an upstream ribosome binding site sequence (optimally  
5-13bp before the initiation codon). If this cannot be identified  
we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Location/Qualifiers  
1..40245  
/organism="Mycobacterium leprae"  
/db\_xref="taxon:1769"  
/clone="cosmid B2533"  
/complement(1..1099)  
/gene="ansp"  
complement(1..31687)  
/misc\_feature  
/Note="Overlap with EMBL:ML017 cosmid B2126 from 1 to  
31682. There are 16 conflicts between this sequence and  
ours. In each case our sequence has been checked and is  
thought to be correct"  
complement(1..1099)  
CDS  
/gene="ansp"  
/Note="MLCB2533.01c, ansp, probable L-asparagine permease,  
partial CDS, len: >366 aa; highly similar to many

amino-acid permeases e.g. ANSP\_SALTY (EMBL:U04851)  
S.typhimurium Ansp, L-asparagine permease (L-asparagine  
transport protein) (497 aa), fasta scores; opt: 1308  
z-score: 1696.8 E(): 0, 61.2% identity in 366 aa overlap.  
Equivalent to M.tuberculosis RV2127, ansp (MTCY261.26,  
85.7% identity in 356 aa overlap). Also similar to  
M.tuberculosis permease RV0346c, aroP2 (MTCY1310.06c,  
75.1% identity in 345 aa overlap). Probable integral  
membrane protein, contains PS00218 Amino acid permease  
signature. Pfam match to entry PF00324 aa\_permeases, amino  
acid permease. Annotated as ORF TR:Q49801, designated  
aroP2 in M.leprae cosmid EMBL:U00017"  
/codon\_start=1  
/transl\_table=11  
/product="putative L-asparagine permease"  
/protein\_id="CAA22915.1"  
/db\_xref="GI:4200259"  
/translation="MATLAESPPEKSGASRAGVLGEAGYHKGKLPQLOMIGIGGAI  
GTGFLGAGRLAKAGPGLFVYAVCGVFVFLIRALGELVLRHPPSSGVSVAAREFF  
GEAAVYVGLYFLDWMATVDTTATITLHRWTFTALPQWTLALLAVLVNML  
ISVWEGELEFMAALIKVCALMAFLVVGTFILGGRYPVDGHTGLSLTSHGGIFPTG  
VALLIVSSGVMFAVAVELVGTAGETVEPKIMPRAINSVIARIAIFYVGSVILLA  
LLPIYSAKASESPFTFFSKVGFYAGDLMNIVLTAALSSNAGLIATGRVMSHIA  
INGSGPKFTARMKNGVPGYGGILLAAVICLG"  
1..1168  
/Note="1168 bp perfect direct repeat"  
complement(2..991)  
/misc\_feature  
/Note="Pfam match to entry PF00324 aa\_permeases, Amino  
acid permease, score 245.80, E-value 6e-70"  
438  
/Note="ansp"  
/Note="ambiguous base T / G"  
complement(833..925)  
/misc\_feature  
/Note="ansp"  
/Note="PS00218 Amino acid permeases signature"  
1223  
/Note="conflict: T is TG EMBL:ML017"  
complement(1249..2766)  
/gene="ansp2"  
complement(1249..2766)  
CDS  
/Note="ansp2"  
/Note="MLCB2533.02c, ansp2, probable L-asparagine  
permease, len: 505 aa; highly similar to many amino-acid  
permeases e.g. ANSP\_SALTY (EMBL:U04851) S.typhimurium  
Ansp, L-asparagine permease (L-asparagine transport  
protein) (497 aa), fasta scores; opt: 1891 z-score: 2218.9  
E(): 0, 58.9% identity in 477 aa overlap. Equivalent to  
M.tuberculosis RV2127 (MTCY261.26, 83.7% identity in 485  
aa overlap). Also similar to M.tuberculosis permease  
RV0346c (MTCY1310.06c, 69.8% identity in 473 aa overlap).  
Probable integral membrane protein, contains PS00218 Amino  
acid permeases signature. Pfam match to entry PF00324  
aa\_permeases, amino acid permease. Annotated as ORF  
TR:Q49802, designated lypS in M.leprae cosmid EMBL:U00017"  
/codon\_start=1  
/transl\_table=11  
/product="putative L-asparagine permease"  
/protein\_id="CAA22916.1"  
/db\_xref="GI:4200260"  
/translation="MATLAESPPEKSGASRAGVLGEAGYHKGKLPQLOMIGIGGAI  
GTGFLGAGRLAKAGPGLFVYAVCGVFVFLIRALGELVLRHPPSSGVSVAAREFF  
GEAAVYVGLYFLDWMATVDTTATITLHRWTFTALPQWTLALLAVLVNML  
ISVWEGELEFMAALIKVCALMAFLVVGTFILGGRYPVDGHTGLSLTSHGGIFPTG  
VALLIVSSGVMFAVAVELVGTAGETVEPKIMPRAINSVIARIAIFYVGSVILLA  
LLPIYSAKASESPFTFFSKVGFYAGDLMNIVLTAALSSNAGLIATGRVMSHIA  
INGSGPKFTARMKNGVPGYGGILLAAVICLG"  
1223  
/Note="ansp2"  
/Note="MLCB2533.01c, ansp, probable L-asparagine permease,  
partial CDS, len: >366 aa; highly similar to many

```

repeat_unit
1668..2835
/Note="1168 bp perfect direct repeat"
misc_feature
2104
/gene="ansp2"
/Note="conflict: C is CT in EMBL:ML017"
misc_feature
complement(2500..2592)
/gene="ansp2"
/Note="P500218 Amino acid permeases signature"
gene
complement(2903..3856)
/Note="MLCB2533.03c"
CDS
complement(2903..3856)
/Note="MLCB2533.03c"
/Note="MLCB2533.03c, hypothetical protein, len: 317 aa;
similar to M.tuberculosis hypothetical protein Rv2125
(MTCV261.21) (EMBL:297559) (292 aa), fasta scores: opt:
1648 z-score: 2382.2 E(): 0, 84.1% identity in 290 aa
overlap. Also some similarity to M.leprae hypothetical
protein TR:Q49847 (29.7% identity in 279 aa overlap).
Annotated as ORF TR:Q49797, hypothetical protein in
M.leprae cosmid EMBL:U00017"
/codon_start=1
/transl_table=11
/product="hypothetical protein MLCB2533.03c"
/protein_id="CAA22917.1"
/db_xref="GI:4200261"
/db_xref="SPTREMBL:Q49797"
/translation="WPPHRAVTRHSSALKPYADSVTLRDGDPDRGALPELHNTVVV
AAFEWNASDASGALEHLNAVHEADPIVEIDDEAYDYOVNRPVIRQVDGVTRELV
WPMKRISICRPGSRNVVMGVPEPNMRWTFCELLIADRLNVDVIVIGALLAD
THTRPVPVSGAAYSPESARFEGLEETRYEGTGIAGVQDACVAARIPAVNMFVAAP
HYVHPNPKATVALLRVEDYDVEPLADLPTQAEWDEQAITEIAAEDDELAIEYVH
SLEQKDAEDVDNDALDKIDGDAALAEFERYLRRRRPGR"
4004..7555
/gene="meth"
CDS
4004..7555
/gene="meth"
/Note="MLCB2533.04, meth, probable
5-methyltetrahydrofolate-homocysteine methyltransferase,
len: 1183 aa; similar to many members of vitamin-B12
dependent methionine synthase family e.g. METH_ECOLI
(EMBL:X16584) E.coli meth (1226 aa), fasta scores: opt:
1617 z-score: 1000.7 E(): 0, 31.6% identity in 1228 aa
overlap. Equivalent to M.tuberculosis Rv2124c
(MTCV261.20c, 88.7% identity in 1183 aa overlap).
Annotated as METH_MYCLE, designated meth2 in M.leprae"
alignment_scores
Quality: 44.00 Length: 12
Ratio: 4.400 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 75.000
alignment_block
US-08-653-294-36 x MLCB2533
Align seg 1/1 to: MLCB2533 from: 1 to: 40245
1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
17534 TAUCGATGGCAATCGCAGNATATCGCAGCGTAT 17569
seq_name: gb_bal:U00017
seq_documentation_block:
LOCUS U00017 42157 bp DNA BCT 01-MAR-1994
DEFINITION Mycobacterium leprae cosmid B2126.
ACCESSION U00017
VERSION U00017.1 GI:466994
KEYWORDS
SOURCE
ORGANISM Mycobacterium leprae.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE
1 (bases 1 to 42157)
Smith,D.R.
Unpublished
2 (bases 1 to 42157)
Robison,K.
Direct Submission
Submitted (01-NOV-1993) Department of Genetics, Harvard Medical
School, 200 Longwood Avenue, Boston MA 02115
3 (bases 1 to 42157)
Robison,K.
Direct Submission
Submitted (01-MAR-1994) Department of Genetics, Harvard Medical
School, 200 Longwood Avenue, Boston MA 02115
On Mar 31, 1994 this sequence version replaced gi:414223.
This sequence data was produced by the Genome Sequencing Center
located at Collaborative Research Incorporated (1365 Main St.,
Waltham MA, 02159). 617-487-7979). Please contact Doug Smith
(smith@cr.cric.com). The annotation should be considered
preliminary and incomplete.
Location/Qualifiers
1. 42157
/organism="Mycobacterium leprae"
/db_xref="taxon:1769"
complement(133..1086)
/Note="match to yigU and yigV E.coli; B2126_C1_183"
/codon_start=1
/transl_table=11
/product="u2126a"
/protein_id="AAAL7191.1"
/db_xref="GI:467006"
/translation="MNGCRVAVRACDLKRIKQHYRRSTNPDMATSLIDHLTELRT
LLISLAAVVTITFGFIWYSHSIFGSLGELWRPQYSLQPSARADISPDQCRLLA
TAPDOFLRIKVGMAAGIVLASPMFYQLWAFITPGLYTKERFTVAFAVPAALVFA
GGTVLAVLQKALGELLIVGSGVQVATSGDRYEGFLNLLVVGVSFEPLLVML
NIAGLLTYRLKSNRRGLIFAMFPAVFTPGSDPFSMTALGALTVLLELAIQLVRL
HDKRRVKHEALIDAEASVIEPPSSIPERSYATSTRSHDDVT"
complement(1132..1398)
/Note="match to yigT E.coli; B2126_C1_182"
/codon_start=1
/transl_table=11
/product="u2126b"
/protein_id="AAAL7190.1"
/db_xref="GI:467005"
/translation="MGLSPWHVVLVWVLLFGAKKLPDAARSLGKSMRIFKSEL
EMOTENQAASALETPMONTPTVQSQRVVPPVWSTEQDTEARPA"
complement(1471..1692)
/codon_start=1
/transl_table=11
/product="B2126_C1_181"
/protein_id="AAAL7189.1"
/db_xref="GI:467004"
/translation="MSWIFEVYPMROAHELPGDSFOAVMTYASDAWMTRLVLGFS
QVAPALAYVRNNAVALESYQVTAQA"
complement(1701..2444)
/codon_start=1
/transl_table=11
/product="B2126_C2_220"
/protein_id="AAAL7199.1"
/db_xref="GI:467014"
/translation="MTQSLRVLRLNNVYFQANPRITRAEAADLVGSAKQLDDF
NQLWVGLPGYGGDLIDFEFSGDTIEVTSAGIDRPLQLTSPEAIGLLVALRLANI
PGVVDPEAVRSAGIAIEAAVVMGNEATGVSVDTRFSESHVAARAVRNRKQAL
VIDYTSASHDLTSLRVDPILVLLVGDHSYLEANSREAGVRLFRDRIIVVARELDEP
AAAPTVRRRCOTHRSSMTTCRRRCG"
complement(2441..3436)
/codon_start=1
/transl_table=11
/product="B2126_C3_266"
/protein_id="AAAL7208.1"
/db_xref="GI:467023"
/translation="MATSKVRLNVLVALLSTRGYTAEKIRSSVAGSYSDPTVEAF
SRMFERNELRDGIPLEVKVSALDPSEGRINRDYALPPPELTDPDEAAVAVAT
OLWESQELITATQGALLKRAAGVDIDPDTVPVVIASSSGVSSLRGSEDFLSILSAI

```

```

GSRQAVQFYPYRPSRAEPTVMRNVPEWGVITENSCWLVGHDCDRNATRTFLRSIGSE
VAPIGPAGVTPVPGVLDRLRVSDAIEVSTGATARYVVDGATALRHAGRPAGVER
LGGRDQVIELDGSIDRLARDIAGHGADAVLEPDLRDDVLIRLRAHAGTGPS"
complement(3443. .3841)
/codon_start=1
/transl_table=11
/product="B2126_C3_265"
/protein_id="AA17207.1"
/db_xref="GI:467022"
/translation="MDTEIDWVIRKRLFRQYQDRNNMELDPKIAQLDLAYHDIKRRG
GVFDLLQKGLAARTDEDIADAVNHPPTTRARLRGEFISAQAAGRDFTDVWVHL
KLNDQAQRTVLCKDPFRAVDVERVKRLIASM"
complement(3861. .4871)
/codon_start=1
/transl_table=11
/product="B2126_C2_219"
/protein_id="AA17198.1"
/db_xref="GI:467013"
/translation="MVGWYPMWLAHRDAOTTSTLKDVRIMGIETEFVGTCTFHGH
RRLSPDEVARYLFRVSVSGRSSNVFLRNGARLYLDVSGSHPEVATAECDNLVOLVTHD
RAGWVEDLLYDAEQRLADEIGGDIYLLKNNYDSAGNSYGGHENLYIVRAGEFSRI
SDVLLPFLYQLCAGKVLQTPKAAATFCLSQRAEHIEGVSATIRPIINIRDE
PHADAERYRLHIVGDSNMCETTMLKVGTAALMEVETGPTGFDLSLDNPIRAIR
EVSHDITGRPRVLRAGGRQASALDIQREYITRAFEHLQTPREPNVQFEQVYDLWGPSAR
RH"
complement(6569. .6862)
/codon_start=1
/transl_table=11
/product="B2126_C3_261"
/protein_id="AA17206.1"
/db_xref="GI:467021"
/translation="MKIADANVLLYSGNTSSEHTGHPGAGSTVRLSCADRGIGWVPL
LVFVRPAKMGVLVLRMTSSSEDAIQGVADNLGTGSAVLMCLTVRHAELVKILY"
complement(6822. .7115)
/codon_start=1
/transl_table=11
/product="B2126_C2_217"
/protein_id="AA17197.1"
/db_xref="GI:467012"
/translation="MSPKKVFNDVIRSVLGRPEHPHFTACITDLSLAVNLDWALPI
AAEFDEELVRLUGFGHERRSFLSHRLAPHLFARIITIHENRRRRAAVER"
complement(7253. .8050)
/codon_start=1
/transl_table=11
/product="B2126_C3_260"
/protein_id="AA17205.1"
/db_xref="GI:467020"
/translation="MSPFYFISPEQAMREQRELARKGIARSGSVLALYAGGVLFVAE
NPSRLQKISELYDRVFAAGKFNEDNLRGGIQPADTRGYAYDRDYTRQLANV
YAOTLGTIFTEQAKPYEVELCVAEVHAYGETKPELYRITYDGSINDEPHFWMGGTT
ESTANALKESYAENASLTDALGIATAALRAGSADAAGSDOPTLGVASLEVAVLDANRP
RRAPRILIGSGLEALLREKSKSGKQNGKAGDSKNSKSYGESTD"
complement(8047. .8922)
/codon_start=1
/notes="proteasome, beta subunit; B2126_C1_173"
/transl_table=11
/product="prcb"
/protein_id="AA17188.1"
/db_xref="GI:467003"
/translation="MTRSFDPRLPTNLAFPGISVINOSFVLLRRQAPELLVPSLGG
GGGGQQLSGHTTIVLYKYPGGVYIAGDRSRSTQGNMAGRDVRKVIYTDYATATGIA
GIAAVAYEFARLVALEHVEKLEGVPLTEPAGKVNRLAIMVRSNLTAAAGLGLALPLL
AGYDIHAPDQSGAGRIYSFDAAGWNTEEGYQSGSGSIFAKSSIKKLYSOYSDADS
ALRVATCALYDAADDSDATGGPDLVRGIYTYNTIGABGAETVESIAELAREIIES
RSRYITLUGSGGSEK"
complement(9142. .10686)
/codon_start=1
/transl_table=11
/product="B2126_C1_172"
/protein_id="AA17187.1"
/db_xref="GI:467002"
/translation="MQRIITEVEYIGSSPDPNPANFILSTQAVLAYAAAAQIQRWK
RTRWDYEVESPLRDARGFDLSRAGSPPPVVDADVGAANNILTNGLRXYVDHAPPEYS
```

```

APBCTDPLDAIWDKAGERVMAAARHVASVPGAACKLQLYKNNVDGKGSYGAHENYL
MSQTPSPSIIAGLTPLVSRVVTGSGRVGIGPAGDEPGFOLQSORSDDYLEVEVGLET
TLKRGINTREDEPHADADRYRLRHVIVGDNANLAETSTYKLTALTALVDLIEGPPVHG
IDTDDTLARPVHAVHAIRODASLRATVTLVDGRELTLGALQRIYLDRAKLVDSRDP
DPRADVVTWVHVLDERDPMDCAEILLDMPAKRLLEGFRQRENLNWSAPRLHLVD
LOYSDVRLDKGLYNRLVARGSMKRLNEHQLRAVNNPPTDTRAYFRGECLRFSSADI
AAASWDSVIFDLGGDSLVRITPLEPLRGSKAHVGLLSDSVDSAAELVEQLTKRPVDPG
KVGLDR"
/loop19. .11611
/rpt_family="RLEP"
complement(11986. .12465)
/codon_start=1
/transl_table=11
/product="B2126_C1_169"
/protein_id="AA17186.1"
/db_xref="GI:467001"
/translation="MSSALTYLARFLLMKKEEYVAELDDSYADYLDLNCYKPDV
QMFESVLNIFVAARPSAKTAVIAPVILKGPFTGTGAPALDGAKSTFKHDDSWTDEM
AHGRSLDAUTLQPRHRLQFNIRISDLECEDSRITAAALISTDSRRLACDYSLPALSDAA"
complement(12950. .13234)
/codon_start=1
/transl_table=11
/product="B2126_C3_259"
/protein_id="AA17204.1"
/db_xref="GI:467019"
/translation="MAVGDDDDRRHSAGLLSSIDRDHWGLQPOSSPAASTVQPLPASR
IPKTSFKPYHVMMLGSRGASSLTALVQAALDPVKQNCFRSRVLFGT"
complement(13405. .13665)
/transl_table=11
/product="B2126_F1_17"
/protein_id="AA17209.1"
/db_xref="GI:467024"
/translation="MVIVKSERVADLLSVLVVARNVDLGYQRFGRAGMGDANDVT
PDSRCPPGPNPPVLLTLVFNVTYVWVATTLIREHYGR"
complement(13640. .13921)
/codon_start=1
/transl_table=11
/product="B2126_C3_258"
/protein_id="AA17203.1"
/db_xref="GI:467018"
/translation="MVAVTATYICISCEVAITKTLRRSALVLSAALGEASSIAATG
SVRGPAFRSSYSGSTVDFIVGLTRFDGGEDTGAFAPVVRKIIASHAP"
complement(14353. .14763)
alignment_scores:
Quality: 44.00 Length: 12
Ratio: 4.400 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 75.000
alignment_block:
US-08-653-294-36 x U00017/rev ..
Align seg 1/1 to reverse of: U00017 from: 1 to: 42157
1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
||||| 14153 TATCGATTGGCAAGTCGCAGATATCTCGACGGTAT 14118
seq_name: gb_ba2:AF026541
seq_documentation_block:
LOCUS AF026541 1692 bp DNA BCT 30-OCT-1998
DEFINITION Mycobacterium tuberculosis Ceob (ceob) gene, complete cds; and Ceoc
(cceoc) gene, partial cds.
ACCESSION AF026541
VERSION AF026541.1 GI:2582553
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
Bacteri; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
```

```

REFERENCE
AUTHORS      1 (bases 1 to 1692)
TITLE        Novel selection for isoniazid (INH) resistance genes supports a
              role for NAD+-binding proteins in mycobacterial INH resistance
JOURNAL      Infect. Immun. 66 (11), 5099-5106 (1998)
MEDLINE      99003115
REFERENCE
AUTHORS      2 (bases 1 to 1692)
TITLE        Direct Submission
JOURNAL      Submitted (23-SEP-1997) Molecular Microbiology and Immunology,
              Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205,
              USA
FEATURES
Source       Location/Qualifiers
              1. .1692
                /organism="Mycobacterium tuberculosis"
                /strain="H37Rv"
                /db_xref="taxon:1773"
              681. .1364
                /gene="ceob"
              681. .1364
                /gene="ceob"
                /note="trkA homolog; complements the INH-sensitive
                phenotype of oxyR deletion mutant of E. coli"
                /codon_start=1
                /transl_table=11
                /product="Ceob"
                /protein_id="AAC69359.1"
                /db_xref="GI:2582554"
                /translation="MRVVMGCGRGVAGVADGLSRIGHEVAIDRDSAFNRLSPQFA
                GERVLGGQDRDLRLRAGIQGADAAVSSGDSNLIISARLARETFGVRVARIYDA
                KRAEYERLIGPTITVTPTDRLNALMODTETAKWRDPTGTVAEVLHEDVWGH
                RATDLEQATGARIAFLIREGTGVLPEKTVLQAGDKVYIAAISGRAEAAALALPLPS
                EDFEGARR"
              1361. >1692
                /gene="ceoc"
              1361. .1692
                /gene="ceoc"
                /note="trkA homolog; complements the INH-sensitive
                phenotype of oxyR deletion mutant of E. coli"
                /codon_start=1
                /transl_table=11
                /product="Ceoc"
                /protein_id="AAC69360.1"
                /db_xref="GI:2582555"
                /translation="MKVAVAGAGVGRSVTRELVENGHDTLIERNPHDLAAAIPEA
                HWRLGDGCLSLLESJLHLEEDFVVAATGDDRVNVVLSLAKTEAFVPRVAVRNDPR
                NEWLPNDA"
BASE COUNT   337 a 544 c 550 g 261 t
ORIGIN

alignment_scores
  Quality: 42.00      Length: 12
  Ratio: 4.200       Gaps: 0
  Percent Similarity: 83.333 Percent Identity: 75.000

alignment_block
US-08-653-294-36 x AF026541 ..
Align seg 1/1 to: AF026541 from: 1 to: 1692
1 TyrArgLeuAlaIleArgArgIleAlaIleuArgTyr 12
||||| ||||||| |||:|||||||
1252 TACAGCGCGGATAGCTATATCGTCGCATAT 1287

seq_name: gb_bal:MTCY05A6
seq_documentation_block:
LOCUS      MTCY05A6 38631 bp DNA BCT 17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 120/162.
ACCESSION 296072 ALI23456
VERSION   296072.1 GI:3261793
KEYWORDS

```

```

SOURCE
ORGANISM      Mycobacterium tuberculosis.
              Mycobacterium tuberculosis
              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Corynebacterineae; Mycobacteriaceae;
              Mycobacterium.
REFERENCE
AUTHORS      1 (bases 1 to 38631)
              Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
              Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
              Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
              Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
              Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
              Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
              Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
              Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
              Barrell, B.G.
              Deciphering the biology of Mycobacterium tuberculosis from the
              complete genome sequence
              Nature 393 (6685), 537-544 (1998)
              98295987
              Erratum: [[published erratum appears in Nature 1998 Nov
              12:396(6707):190]]
              2 (bases 1 to 38631)
              Parkhill, J.
              Direct Submission
              Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
              tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
              Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
              Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
              On Jun 27, 1998 this sequence version replaced gi:2181970.
              Notes:
              Details of M. tuberculosis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have
              been renumbered from the original cosmid submissions but the old
              gene designations are in brackets after the new gene numbers.
              Gene prediction was based on a Hidden Markov Model of TB genes
              implemented in TBparse (Krogh) supplemented with visual inspection
              of positional base preference in codons, especially where there is
              an increase in the observed/expected third position G + C.
              CAUTION: In some cases we may not have predicted the correct
              initiation codon. Where possible we choose an initiation codon
              (atg, gtg, or ttg) which is preceded by an upstream ribosome
              binding site sequence (optimally 5-13bp before the initiation
              codon). If this cannot be identified we choose the most upstream
              initiation codon.
FEATURES
source       Location/Qualifiers
              1. .38631
                /organism="Mycobacterium tuberculosis"
                /strain="H37Rv"
                /db_xref="taxon:1773"
              <1. .763
                /note="fragment designated v010. Does not represent a
                physical clone"
              102. .734
                /gene="Rv2680"
              102. .734
                /gene="Rv2680"
                /note="Rv2680" (MTV010.04), len: 210. Unknown but very
                similar to M. leprae protein U1764V. FASTA scores:
                sptr|Q50005|Q50005 U1764V (213 aa) opt: 1136 z-score:
                1231.1 E(): 0.83.9% identity in 193 aa overlap. TBparse
                score is 0.886"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2680"
                /protein_id="CAB09496.1"
                /db_xref="GI:3261794"
                /db_xref="SPTREMBL:O86317"
                /translation="MTSAGDAAERSDDEERLRTSAEPALFREAVAAAMVAVRPEIEL
                GPIRPPQLAPYSYALGAEIKHPELDVIPERSEGDAGFLNMLYDPDGSADWDGIRL
                VAYVQADLDSAEVDPDLPEVAWSWLVDALTARQVRLGVTGTTATTSVRYGDISGP
                PRAHOLELRASWTATTDLGAHVQAFCDVLEHAAGLPAGVPAGVTDLGRSRA"

```



```

SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      1 (bases 1 to 142796)
TITLE        Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL      Homo sapiens, clone RP11-7024
REFERENCE
AUTHORS      2 (bases 1 to 142796)
TITLE        Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Castelle,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Kleh,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL      Direct Submission
COMMENT      Submitted (15-NOV-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 10, 1999 this sequence version replaced gi:6425750.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTHR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L2959
Center clone name: 7_O_24
-----
Summary Statistics
Sequencing vector: MJ3; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 128649 bases at least Q40
Consensus quality: 137112 bases at least Q30
Consensus quality: 140758 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 142796; sum-of-contigs
Quality coverage: 5.7 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2142: contig of 2142 bp in length
*      gap of unknown length
* 2143 4834: contig of 2692 bp in length
*      gap of unknown length
* 4835 9211: contig of 4377 bp in length
*      gap of unknown length
* 9212 25157: contig of 15946 bp in length
*      gap of unknown length
* 25158 66208: contig of 41051 bp in length
*      gap of unknown length
* 66209 142796: contig of 76588 bp in length.
*      Location/Qualifiers
*      1. .142796
*      /organism="Homo sapiens"
*      /db_xref="taxon:9606"

```

```

/clone="RP11-7024"
/clone_lib="RPCI-11 Human Male BAC"
BASE COUNT 47989 a 24580 c 24540 g 45686 t 1 others
ORIGIN

alignment_scores:
  Quality: 42.00      Length: 12
  Ratio: 4.200       Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667

alignment_block:
US-08-653-294-36 x AC013772/rev ..
Align seg 1/1 to reverse of: AC013772 from: 1 to: 142796
1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13521 TATAGACTAGCAGTCAGAGACATTTTACTTAATAT 13486

seq_name: gb_htg5:AC013205

seq_documentation_block:
LOCUS      AC013205      56237 bp      DNA      HTG      03-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION  AC013205
VERSION    AC013205.1 GI:6223127
KEYWORDS  HTG; HTGS_PHASE2.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 56237)
AUTHORS   Adams,M. and Venter,J.C.
TITLE     Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10214117 by the submitter.
For further information on this sequence you may e-mail to
fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
*      Location/Qualifiers
*      1. .56237
*      /organism="Drosophila melanogaster"
*      /db_xref="taxon:7227"
BASE COUNT 16357 a 12291 c 11944 g 15645 t
ORIGIN

alignment_scores:
  Quality: 40.00      Length: 11
  Ratio: 4.000       Gaps: 0
Percent Similarity: 90.909 Percent Identity: 72.727

alignment_block:
US-08-653-294-36 x AC013205 ..
Align seg 1/1 to: AC013205 from: 1 to: 56237
2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
47041 CGACTGGCAGTCGACGTACTACGTCGATAC 47073

seq_name: gb_bal:MTV002

seq_documentation_block:
LOCUS      MTV002      56414 bp      DNA      BCT      17-JUN-1998.
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.

```

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REMARK	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
AL008967	AL123456		GI:3261491	Mycobacterium tuberculosis.												
				Mycobacterium tuberculosis.												
				Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.												
				1 (bases 1 to 56414)												
				Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmair, K., Gas, S., Barry III, C.E., Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrall, B.G.												
				Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence												
				Nature 393 (6685), 537-544 (1998)												
				98295987												
				Erratum: [[published erratum appears in Nature 1998 Nov 12; 396(6707):1901]]												
				2 (bases 1 to 56414)												
				Parkhill, J.												
				Direct Submission												
				Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk												
				On Jun 27, 1998 this sequence version replaced gi:2624256.												
				Notes:												
				Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.												
				(URL, <a href="http://www.sanger.ac.uk/projects/M_tuberculosis/">http://www.sanger.ac.uk/projects/M_tuberculosis/</a> ) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.												
				Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.												
				CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.												
				FEATURES												
				Source												
				1, 56414												
				/organism="Mycobacterium tuberculosis"												
				/strain="H37Rv"												
				/db_xref="taxon:1773"												
				<1. >56414												
				/note="fragment designated v002. Does not represent a physical clone"												
				<1. .233												
				/organism="Mycobacterium tuberculosis"												
				/strain="H37Rv"												
				/db_xref="taxon:1773"												
				/clone="Y154"												
				complement(3. .527)												
				/gene="recX"												
				complement(3. .527)												
				/gene="recX"												
				/note="RV2736c, (MTV002.01c), len: 174 aa. recX, similar to eg. RECX_PSEAE_P37860 regulatory protein recX from Pseudomonas aeruginosa (153 aa), fasta scores; opt: 161 z-score: 257.2 E(1): 3.6e-07, 30.7% identity in 137 aa overlap. Overlaps and extends CDS from overlapping cosmid MTCY154.16c"												
				<1. .233												
				/organism="Mycobacterium tuberculosis"												
				/strain="H37Rv"												
				/db_xref="taxon												



```

/translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHLVLTGDFGRSAAQAI
EDGVEPRDVNRALCADFDVPHDRW"
gene complement(3464..4630)
CDS /gene="Rv2739c"
complement(3464..4630)
/gene="Rv2739c"
/translation="MTV002.04c", len: 388 aa. Probable
transferase, similar to eg. TR:Q51560 (EMBL:L28170)
Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
identity in 425 aa overlap. Equivalent to Mycobacterium
leprae protein MCB33.02c (392 aa); fasta scores
gpi294723|MCB33_2 Mycobacterium leprae cosmid B33 opt:
2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2739c"
/protein_id="CAAI5535.1"
/db_xref="GI:2624261"
/db_xref="SPTREMBL:O33282"
/translation="MRVAVVAGDPGHSFPAIALCORFAAADTPILTGVWLEAAR
AAGDAVELDGLAATDRLDAGAKIHRRAQMAVNVPRRLALEPELVVSDVITACGG
MAAELGLGVNPHDPLPGLPIGSLAAGTIRGLRDATMRALTGRSWRAG
LRQAAVRVVGIPARDPGLRLIATLPALVPRDPWPAPAEVAVVVGSLFEDTDRVLA
IPAGTGVVVVASTALTGTAGLTVLQSLTGTCTVPSGLRVSLSGADLTVPWP
AVAGLSQAEALLRADLVICGGGGHGWAKTLLAGVPMVAVVPGGDDQWEIANRVVROGS
AVILRPTDALVAANVEVLSSRFREARRAAASVAGAADPRVCHDALALAG"
gene 4674..5123
CDS /gene="Rv2740"
4674..5123
/gene="Rv2740"
/translation="Rv2740", len: 149 aa. unknown, similar
to M. leprae hypothetical protein TR:Q49850 (EMBL:Z94723)
MCB33.03 (B2235_F3_140) (178 aa), fasta scores: opt: 498
z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2740"
/protein_id="CAAI5536.1"
/db_xref="GI:2624262"
/db_xref="SPTREMBL:O33283"
/translation="MAELTSTPETTEAIRAVEAFLNALQNEFDITVDAALGDDL
VYENVGSRIRGRRRTATLLRMQGRVGFVKIHRIGADGAATVTERDAILIPLRV
OFWVCGVEFDGGRITLWRDYFDVDMFKLLRLVALVPSLKATL"
gene 5355..6932
CDS /gene="PE_PGRS"
5355..6932
/gene="PE_PGRS"
/translation="Rv2741", len: 525 aa; Member of M.
tuberculosis PE_PGRS subfamily, similar to many eg.

alignment_scores:
Quality: 40.00 Length: 9
Ratio: 4.444 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-08-653-294-36 x MTV002/rev ..
Align seg 1/1 to reverse of: MTV002 from: 1 to: 56414
1 TyrArgLeuAlaIleArgArgIleAla 9
|||||:|||||:|||||:|||||:|||||:
35552 TACCGGATCGCCATTCGCAGATCGCG 35526

seq_name: gb_htg7:AC017383

seq_documentation_block:
LOCUS AC017383 115873 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
pieces.

/translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHLVLTGDFGRSAAQAI
EDGVEPRDVNRALCADFDVPHDRW"
gene complement(3464..4630)
CDS /gene="Rv2739c"
complement(3464..4630)
/gene="Rv2739c"
/translation="MTV002.04c", len: 388 aa. Probable
transferase, similar to eg. TR:Q51560 (EMBL:L28170)
Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
identity in 425 aa overlap. Equivalent to Mycobacterium
leprae protein MCB33.02c (392 aa); fasta scores
gpi294723|MCB33_2 Mycobacterium leprae cosmid B33 opt:
2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2739c"
/protein_id="CAAI5535.1"
/db_xref="GI:2624261"
/db_xref="SPTREMBL:O33282"
/translation="MRVAVVAGDPGHSFPAIALCORFAAADTPILTGVWLEAAR
AAGDAVELDGLAATDRLDAGAKIHRRAQMAVNVPRRLALEPELVVSDVITACGG
MAAELGLGVNPHDPLPGLPIGSLAAGTIRGLRDATMRALTGRSWRAG
LRQAAVRVVGIPARDPGLRLIATLPALVPRDPWPAPAEVAVVVGSLFEDTDRVLA
IPAGTGVVVVASTALTGTAGLTVLQSLTGTCTVPSGLRVSLSGADLTVPWP
AVAGLSQAEALLRADLVICGGGGHGWAKTLLAGVPMVAVVPGGDDQWEIANRVVROGS
AVILRPTDALVAANVEVLSSRFREARRAAASVAGAADPRVCHDALALAG"
gene 4674..5123
CDS /gene="Rv2740"
4674..5123
/gene="Rv2740"
/translation="Rv2740", len: 149 aa. unknown, similar
to M. leprae hypothetical protein TR:Q49850 (EMBL:Z94723)
MCB33.03 (B2235_F3_140) (178 aa), fasta scores: opt: 498
z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2740"
/protein_id="CAAI5536.1"
/db_xref="GI:2624262"
/db_xref="SPTREMBL:O33283"
/translation="MAELTSTPETTEAIRAVEAFLNALQNEFDITVDAALGDDL
VYENVGSRIRGRRRTATLLRMQGRVGFVKIHRIGADGAATVTERDAILIPLRV
OFWVCGVEFDGGRITLWRDYFDVDMFKLLRLVALVPSLKATL"
gene 5355..6932
CDS /gene="PE_PGRS"
5355..6932
/gene="PE_PGRS"
/translation="Rv2741", len: 525 aa; Member of M.
tuberculosis PE_PGRS subfamily, similar to many eg.

alignment_scores:
Quality: 40.00 Length: 9
Ratio: 4.444 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-08-653-294-36 x MTV002/rev ..
Align seg 1/1 to reverse of: MTV002 from: 1 to: 56414
1 TyrArgLeuAlaIleArgArgIleAla 9
|||||:|||||:|||||:|||||:|||||:
35552 TACCGGATCGCCATTCGCAGATCGCG 35526

seq_name: gb_htg7:AC017383

seq_documentation_block:
LOCUS AC017383 115873 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
pieces.

```

```

AC017383
AC017383.1 GI:6553603
HTG: HTGS_PHASE2.
fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 115873)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210513 by the submitter.
* For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
Source 1..115873
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 31582 a 25315 c 25855 g 33121 t
ORIGIN
alignment_scores:
Quality: 40.00 Length: 12
Ratio: 4.444 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 66.667
alignment_block:
US-08-653-294-36 x AC017383
Align seg 1/1 to: AC017383 from: 1 to: 115873
1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||:
105058 TATTTGCTAGTCGCGCGGCGTGTGCTTAC 105093
seq_name: gb_htg2:AC007417
seq_documentation_block:
LOCUS AC007417 125448 bp DNA HTG 02-AUG-1999
DEFINITION Drosophila melanogaster chromosome 2 clone BACR48F07 (D625) RPCI-98
48.F.7 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 94 unordered pieces.
AC007417
AC007417.3 GI:5670592
HTG: HTGS_PHASE1.
fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 125448)
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.N., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zier, L.L. and
Rubin, G.M.
TITLE Sequencing of Drosophila melanogaster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 125448)
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.N., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,

```

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.

## TITLE

Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

## COMMENT

On Aug 2, 1999 this sequence version replaced gi:5629944.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 94 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 970 1049: contig of 969 bp in length  
\* 1050 2078: contig of 1029 bp in length  
\* 2079 2158: gap of unknown length  
\* 2159 3249: contig of 1091 bp in length  
\* 3250 3259: gap of unknown length  
\* 3330 4547: contig of 1218 bp in length  
\* 4548 4627: gap of unknown length  
\* 4628 5262: contig of 635 bp in length  
\* 5263 5342: gap of unknown length  
\* 5343 6096: contig of 754 bp in length  
\* 6097 6176: gap of unknown length  
\* 6177 7486: contig of 1310 bp in length  
\* 7487 7566: gap of unknown length  
\* 7567 8160: contig of 594 bp in length  
\* 8161 8240: gap of unknown length  
\* 8241 9090: contig of 850 bp in length  
\* 9091 9170: gap of unknown length  
\* 9171 9803: contig of 733 bp in length  
\* 9804 9983: gap of unknown length  
\* 9984 10564: contig of 581 bp in length  
\* 10565 10644: gap of unknown length  
\* 10645 11655: contig of 1011 bp in length  
\* 11656 11735: gap of unknown length  
\* 11736 12459: contig of 724 bp in length  
\* 12460 12339: gap of unknown length  
\* 12340 13197: contig of 658 bp in length  
\* 13198 13277: gap of unknown length  
\* 13278 14314: contig of 1037 bp in length  
\* 14315 14394: gap of unknown length  
\* 14395 15048: contig of 654 bp in length  
\* 15049 15128: gap of unknown length  
\* 15129 16091: contig of 963 bp in length  
\* 16092 16171: gap of unknown length  
\* 16172 16817: contig of 646 bp in length  
\* 16818 16897: gap of unknown length  
\* 16898 17598: contig of 701 bp in length  
\* 17599 17678: gap of unknown length  
\* 17679 18750: contig of 1072 bp in length  
\* 18751 18830: gap of unknown length  
\* 18831 19376: contig of 546 bp in length  
\* 19377 19456: gap of unknown length  
\* 19457 20255: contig of 799 bp in length  
\* 20256 20335: gap of unknown length  
\* 20336 20957: contig of 622 bp in length  
\* 20958 21037: gap of unknown length  
\* 21038 22187: contig of 1150 bp in length  
\* 22188 22267: gap of unknown length  
\* 22268 22871: contig of 604 bp in length  
\* 22872 22951: gap of unknown length  
\* 22952 24044: contig of 1093 bp in length

24045 24124: gap of unknown length  
\* 24125 25189: contig of 1065 bp in length  
\* 25190 25269: gap of unknown length  
\* 25270 26217: contig of 948 bp in length  
\* 26218 26297: gap of unknown length  
\* 26298 27312: contig of 1015 bp in length  
\* 27313 27392: gap of unknown length  
\* 27393 28934: contig of 1542 bp in length  
\* 28935 29014: gap of unknown length  
\* 29015 29998: contig of 984 bp in length  
\* 29999 30078: gap of unknown length  
\* 30079 31456: contig of 1378 bp in length  
\* 31457 31536: gap of unknown length  
\* 31537 32424: contig of 1706 bp in length  
\* 32425 33222: gap of unknown length  
\* 33223 34819: contig of 1497 bp in length  
\* 34820 34899: gap of unknown length  
\* 34900 35765: contig of 866 bp in length  
\* 35766 35845: gap of unknown length  
\* 35846 37032: contig of 1187 bp in length  
\* 37033 37112: gap of unknown length  
\* 37113 38879: contig of 1767 bp in length  
\* 38880 38959: gap of unknown length  
\* 38960 39889: contig of 930 bp in length  
\* 39890 39969: gap of unknown length  
\* 39970 40967: contig of 998 bp in length  
\* 40968 41047: gap of unknown length  
\* 41048 42556: contig of 1609 bp in length  
\* 42557 42737: gap of unknown length  
\* 42738 44172: contig of 1436 bp in length  
\* 44173 44252: gap of unknown length  
\* 44253 45669: contig of 1417 bp in length  
\* 45670 45749: gap of unknown length  
\* 45750 47740: contig of 1991 bp in length  
\* 47741 48827: gap of unknown length  
\* 48828 48907: gap of unknown length  
\* 48908 49770: contig of 863 bp in length  
\* 49771 49850: gap of unknown length  
\* 49851 51951: contig of 2101 bp in length  
\* 51952 52031: gap of unknown length  
\* 52032 53654: contig of 1623 bp in length  
\* 53655 53734: gap of unknown length  
\* 53735 55071: contig of 1337 bp in length  
\* 55072 55151: gap of unknown length  
\* 55152 56770: contig of 1619 bp in length  
\* 56771 56850: gap of unknown length  
\* 56851 58743: contig of 1893 bp in length  
\* 58744 58824: gap of unknown length  
\* 58825 60291: contig of 1468 bp in length  
\* 60292 60371: gap of unknown length  
\* 60372 63759: contig of 3388 bp in length  
\* 63760 63839: gap of unknown length  
\* 63840 68476: contig of 2637 bp in length  
\* 68477 68556: gap of unknown length  
\* 68557 69321: contig of 2765 bp in length  
\* 69322 69401: gap of unknown length  
\* 69402 70806: contig of 1405 bp in length  
\* 70807 70886: gap of unknown length  
\* 70887 73742: contig of 2856 bp in length  
\* 73743 73822: gap of unknown length  
\* 73823 75075: contig of 1253 bp in length  
\* 75076 75155: gap of unknown length  
\* 75156 77979: contig of 2824 bp in length  
\* 77980 78059: gap of unknown length  
\* 78060 81376: contig of 3316 bp in length  
\* 81377 84455: gap of unknown length  
\* 84456 84532: contig of 2997 bp in length  
\* 84533 84533: gap of unknown length  
\* 84534 87596: contig of 3064 bp in length  
\* 87597 87676: gap of unknown length  
\* 87677 102529: contig of 14853 bp in length  
\* 102530 102609: gap of unknown length

AUTHORS Celniker, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,

Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Moschetti, R., Moschetti, M., Nixon, R., Fucied, C.M., Fucied, R.

Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

# Sequencing of Drosophila melanogaster

Unpublished

2 (bases 1 to 183365)

Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleby,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

## Direct Submission

Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Aug 2, 1999 this sequence version replaced g1:5629943.

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [baggs@fruitfly.berkeley.edu](mailto:baggs@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. \* as soon as it is available and the accession number will be preserved.

1 678: contig of 678 bp in length  
 \* 679 758: gap of unknown length  
 \* 759 1693: contig of 935 bp in length  
 \* 1694 1773: gap of unknown length  
 \* 1774 2282: contig of 509 bp in length  
 \* 2283 2362: gap of unknown length  
 \* 2363 2962: contig of 600 bp in length  
 \* 2963 3042: gap of unknown length  
 \* 3043 3789: contig of 747 bp in length  
 \* 3790 3869: gap of unknown length  
 \* 3870 5216: contig of 1347 bp in length  
 \* 5217 5296: gap of unknown length  
 \* 5297 6765: contig of 1469 bp in length  
 \* 6766 6845: gap of unknown length  
 \* 6846 8019: contig of 1174 bp in length  
 \* 8020 8099: gap of unknown length  
 \* 8100 9084: contig of 985 bp in length  
 \* 9085 9164: gap of unknown length  
 \* 9165 9757: contig of 593 bp in length  
 \* 9758 9837: gap of unknown length  
 \* 9838 11177: contig of 1340 bp in length  
 \* 11178 11257: gap of unknown length  
 \* 11258 12992: contig of 1735 bp in length  
 \* 12993 13072: gap of unknown length  
 \* 13073 17311: contig of 4239 bp in length  
 \* 17312 17391: gap of unknown length  
 \* 17392 22798: contig of 5407 bp in length  
 \* 22799 22878: gap of unknown length  
 \* 22879 28458: contig of 5580 bp in length  
 \* 28459 28538: gap of unknown length  
 \* 28539 34268: contig of 5730 bp in length  
 \* 34269 34348: gap of unknown length  
 \* 34349 43284: contig of 8936 bp in length  
 \* 43285 43364: gap of unknown length  
 \* 43365 56700: contig of 13336 bp in length  
 \* 56701 56780: gap of unknown length  
 \* 56781 76800: contig of 20020 bp in length  
 \* 76801 76880: gap of unknown length  
 \* 76881 120727: contig of 43847 bp in length  
 \* 120728 120807: gap of unknown length  
 \* 120808 177030: contig of 56223 bp in length  
 \* 177031 177110: gap of unknown length

\* 177111 177843: contig of 733 bp in length  
 \* 177844 177923: gap of unknown length  
 \* 177924 178557: contig of 634 bp in length  
 \* 178558 178637: gap of unknown length  
 \* 178638 179250: contig of 613 bp in length  
 \* 179251 179330: gap of unknown length  
 \* 179331 179608: contig of 278 bp in length  
 \* 179609 179688: gap of unknown length  
 \* 179689 180406: contig of 718 bp in length  
 \* 180407 180486: gap of unknown length  
 \* 180487 181193: contig of 707 bp in length  
 \* 181194 181273: gap of unknown length  
 \* 181274 181798: contig of 525 bp in length  
 \* 181799 181878: gap of unknown length  
 \* 181879 182592: contig of 714 bp in length  
 \* 182593 182672: gap of unknown length  
 \* 182673 183365: contig of 693 bp in length.

## FEATURES

Source  
 1. 183365  
 /organism="Drosophila melanogaster"  
 /strain="Y: cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="2"  
 /map="47A-47B"  
 /clone\_lib="BACR19J16 (D626) RPCI-98 19.J.16"  
 /clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial ECORI in  
 PBAC83.6"  
 BASE COUNT 51563 a 37925 c 39574 g 51982 t 2321 others  
 ORIGIN

## alignment\_scores:

Quality: 40.00 Length: 12  
 Ratio: 4.444 Gaps: 0  
 Percent Similarity: 75.000 Percent Identity: 66.667

## alignment\_block:

US-08-653-294-36 x AC007352 ..  
 Align seg 1/1 to: AC007352 from: 1 to: 183365  
 1 TyTargLeuAlaIleArgArgIleAlaLeuArgTyr 12  
 ||| ||||| ||||| ||||| ||||| ||||| |||||  
 172452 TATTGCTAGCTGGCGCGCGTGGCTTGCCTAC 172487

OM of: US-08-653-294-36 to: N\_Geneseq\_36: \* out\_format : pfs  
Date: Feb 8, 2000 7:32 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n\_model -DEV=xlp  
-O=/cgl1\_1/USPTO\_SPOOL/US08653294/runat\_04022000\_160701\_15807/app\_query.fasta.2  
-DB=N\_Geneseq\_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

## Search information block:

Query: US-08-653-294-36  
Query length: 12  
Database: N\_Geneseq\_36: \*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 873.190000

score_list:	Seq	Strd	Orig	Zscore	Escore	Len	Documentation
N_Geneseq_36: X13357	+	39.00	110.70	144.91	4114	1	Enterococcus faecalis genome
N_Geneseq_36: Q12225	+	38.00	116.69	67.21	1371	1	Octopus rhodopsin membrane protein
N_Geneseq_36: T28565	+	37.00	117.16	63.21	861	1	Bacterial antibiotic resistance
N_Geneseq_36: N60877	+	37.00	112.94	108.74	1400	1	Sequence encoding Serratia phage
N_Geneseq_36: T26331	+	36.00	120.67	40.34	382	1	Human gene signature HMG508571
N_Geneseq_36: Q99804	+	36.00	115.02	83.23	731	1	PR-1 like gene PR-lm. New DNA
N_Geneseq_36: Q44280	+	36.00	111.53	130.25	1092	1	Pseudomonas cepacia DSM 3401
N_Geneseq_36: T17455	+	36.00	84.63	4.1e+03	24025	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17515	+	36.00	84.63	4.1e+03	24025	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T18325	+	36.00	84.63	4.1e+03	24026	1	BRCA1, human breast and ovar
N_Geneseq_36: T17512	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17513	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17513	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17514	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17516	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17517	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17518	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17519	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17521	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17521	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17522	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17523	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17524	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17526	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17527	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17528	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17529	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17530	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17530	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17532	+	36.00	84.63	4.1e+03	24026	1	BRCA1, human breast and ovar
N_Geneseq_36: T17520	+	36.00	84.63	4.1e+03	24026	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17520	+	36.00	84.63	4.1e+03	24031	1	Mutated BRCA1 genomic sequence
N_Geneseq_36: T17525	+	36.00	84.63	4.1e+03	35100	1	KSHV LTR DNA (nucleotides 1-3
N_Geneseq_36: V73802	+	36.00	81.34	6.2e+03	137507	1	KSHV long unique coding regi
N_Geneseq_36: V19941	+	35.00	69.46	2.8e+04	303	1	Human brain expressed sequen
N_Geneseq_36: Q60295	+	35.00	119.11	49.24	557	1	Truncated p27 protein encoding
N_Geneseq_36: X26229	+	35.00	113.82	97.13	582	1	Human encoding p27 protein. frea
N_Geneseq_36: X17752	+	35.00	113.43	102.01	596	1	Human p27 Kip1 cDNA clone. p27
N_Geneseq_36: T16336	+	35.00	113.23	104.75	597	1	Nucleotide sequence encoding th
N_Geneseq_36: V16719	+	35.00	113.21	104.95	597	1	CDNA encoding protein inhibitin
N_Geneseq_36: V47517	+	35.00	113.21	104.95	597	1	CKI/KIP protein p27 coding sequ
N_Geneseq_36: X21817	+	35.00	113.21	104.95	1028	1	Truncated p27/p16 fusion prote
N_Geneseq_36: X26231	+	35.00	108.48	192.62	1073	1	Truncated p27/p16 fusion prote
N_Geneseq_36: X26232	+	35.00	108.11	201.88	1073	1	CDK inhibitory fusion protein
N_Geneseq_36: T74053	+	35.00	107.91	207.14	1098	1	CDK inhibitory fusion protein

N\_Geneseq\_36: X26224 - 35.00 107.91 207.14 1098 Human p16p27 fusion protein  
N\_Geneseq\_36: X26235 - 35.00 107.73 211.99 1121 Human p16p27 fusion protein  
N\_Geneseq\_36: T74052 - 35.00 107.56 216.63 1143 CDK inhibitory fusion prote  
N\_Geneseq\_36: X26223 - 35.00 107.56 216.63 1143 Human p16(GS)p27 fusion pro

seq\_name: N\_Geneseq\_36: X13357

seq\_documentation\_block:

ID X13357 standard; DNA; 4114 BP.

AC X13357; 1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:420.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN WO9850555-A2.

FD 12-NOV-1998.

PF 04-MAY-1998; U08985.

PR 14-NOV-1997; US-066009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI INC.

FI Barash SC, Dillon FJ, Kunsch CA;

DR WPI; 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

PS Claim 1: Page 1639-1641; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

SQ Sequence 4114 BP; 1334 A; 661 C; 889 G; 1226 T;

alignment\_scores:

Quality: 39.00 Length: 11

Ratio: 3.900 Gaps: 0

Percent Similarity: 90.909 Percent Identity: 63.636

alignment\_block:

US-08-653-294-36 x X13357 ..

Align seg 1/1 to: X13357 from: 1 to: 4114

2 ArgLeuAlaileArgArgileAlaLeuArgTyr 12

|||||:|||||:|||||:|||||:|||||

1110 AGAATGCTTACGACGATTAGCTAGTGCCTAT 1142

seq\_name: N\_Geneseq\_36: Q12225

seq\_documentation\_block:

ID Q12225 standard; DNA; 1371 BP.

AC Q12225; 1991 (first entry)

DE Octopus rhodopsin membrane protein.

KW Octopus; rhodopsin; membrane; helix; OR; ss.

FT Key Location/Qualifiers

FT cds 1..1368

FT /\*tag= a

FT misc\_feature 322..327

FT product= membrane protein

```

FT FT      /*tag= a
FT FT      /label= Sphi
FT FT      /note= "restriction enzyme site"
FT FT      592..597
FT FT      /*tag= b
FT FT      /label= MluI
FT FT      /note= "restriction enzyme site"
FT FT      884..889
FT FT      /*tag= c
FT FT      /label= NdeI
FT FT      /note= "restriction enzyme site"
FT FT      1069..1074
FT FT      /*tag= d
FT FT      /label= NheI
FT FT      /note= "restriction enzyme site"
FT FT      1371
FT FT      /*tag= e
FT FT      /label= BamHI
FT FT      /note= "restriction enzyme site"
FT FT
FT FT      J03123486-A.
FT FT      27-MAY-1991.
FT FT      06-OCT-1989; 260261..
FT FT      08-OCT-1989; JP-260261.
FT FT      (HITA ) HITACHI KK.
FT FT      WPI: 91-197925/27.
FT FT      P-PSDB; R12362.
FT FT      Genetic engineering of membrane protein - by division of protein
FT FT      into cartridge genes corresp. to helix structure polypeptide(s)
FT FT      Disclosure; Fig 1; 17pp; Japanese.
FT FT      A gene cassette is prepd. by division of the OR membrane protein
FT FT      encoding helix structure polypeptides. Restriction sites are indicated.
FT FT      A base sequence contg. such a sequence downstream to the tryptophan
FT FT      regulating gene derived from the E. coli tryptophan operon, the trpL,
FT FT      or the trpE polypeptide translation regulating base sequence, the trpL,
FT FT      and the N-terminal Met of the trpL (or trpE) or trpE polypeptide,
FT FT      respectively, is introduced in an expression vector for
FT FT      transformation of host cells.
FT FT      Sequence 1371 BP; 298 A; 368 C; 351 G; 354 T;
FT FT
FT FT      alignment_scores:
FT FT      Quality: 38.00 Length: 10
FT FT      Ratio: 3.800 Gaps: 0
FT FT      Percent Similarity: 100.000 Percent Identity: 70.000
FT FT
FT FT      alignment_block:
FT FT      US-08-653-294-36 x Q12225/rev ..
FT FT      Align seg 1/1 to reverse of: Q12225 from: 1 to: 1371
FT FT
FT FT      2 ArgLeuAlaIleArgArgIleAlaLeuArg 11
FT FT      1272 CGGTAGCCCTGCGCGGTAGCTGAGG 1243
FT FT
FT FT      seq_name: N_Geneseq_36:T28565
FT FT
FT FT      seq_documentation_block:
FT FT      ID T28565 standard; DNA; 861 BP.
FT FT      AC T28565;
FT FT      DT 01-APR-1997 (first entry)
FT FT      DE Bacterial antibiotic resistance gene, aac2, probe.
FT FT      KW Detection; probe; amplification primer; bacterial pathogen; pneumonia;
FT FT      KW Escherichia coli; Klebsiella pneumoniae; Pseudomonas aeruginosa;
FT FT      KW Proteus mirabilis; Streptococcus pneumoniae; Staphylococcus aureus;
FT FT      KW Staphylococcus epidermidis; Enterococcus faecalis; respiratory tract;
FT FT      KW Staphylococcus saprophyticus; Streptococcus pyogenes; urinary tract;
FT FT      KW Haemophilus influenzae; Moraxella catarrhalis; septicaemia; meningitis;
FT FT      KW infection; intra-abdominal infection; skin infection;
FT FT      KW bacterial resistance; beta-lactam antibiotic; ds.
FT FT      OS Synthetic.
FT FT      PN W09608582-A2.
FT FT      PD 21-MAR-1996.
FT FT      PF 12-SEP-1995; CA0528.

```

---

```

PR PA      12-SEP-1994; US-304732.
PR PA      (BERG/) BERGERON M G.
PR PA      (OUEL/) OUELLETTE M.
PR PA      (ROYP/) ROY P H.
PR PI      Bergeron MG, Ouellette M, Roy PH;
DR WPI: 96-179953/18.
PT Method for the detection of bacterial species using probes and
PT primers - allows detection and quantification of antibiotic
PT resistant bacteria in patients, the environment and food
PS Claim 82; Page 142-143; 216pp; English.
CC The sequences given in T28560-76 represent fragments derived from
CC bacterial antibiotic resistance genes which were used as probes in the
CC method of the invention for the detection of bacterial species in a
CC sample. The method of the invention comprises using probes and/or
CC amplification primers which are specific, ubiquitous and sensitive for
CC determining the presence and/or amount of nucleic acids from selected
CC bacterial species in any sample, where the bacterial nucleic acid
CC comprises a selected target region hybridisable with the probes or
CC primers. The method comprises contacting the sample with the probes or
CC primers and detecting the presence and/or amount of hybridised
CC primers or amplification products as an indication of the presence
CC and/or amount of the bacterial species. This method may be used to
CC detect commonly encountered bacterial pathogens, e.g. Escherichia coli,
CC Klebsiella pneumoniae, Pseudomonas aeruginosa, Proteus mirabilis,
CC Streptococcus pneumoniae, Staphylococcus aureus, Staphylococcus
CC epidermidis, Enterococcus faecalis, Staphylococcus saprophyticus,
CC Streptococcus pyogenes, Haemophilus influenzae and Moraxella catarrhalis.
CC These bacterial species are associated with approx. 90% of urinary tract
CC infections and with a high percentage of other severe infections
CC including septicemia, meningitis, pneumonia, intra-abdominal infections,
CC skin infections and other severe respiratory tract infections. The
CC method may also be used to evaluate a bacterial resistance to beta-
CC lactam antibiotics.
SQ Sequence 861 BP; 174 A; 243 C; 270 G; 174 T;
FT FT
FT FT      alignment_scores:
FT FT      Quality: 37.00 Length: 12
FT FT      Ratio: 3.700 Gaps: 0
FT FT      Percent Similarity: 83.333 Percent Identity: 58.333
FT FT
FT FT      alignment_block:
FT FT      US-08-653-294-36 x T28565 ..
FT FT      Align seg 1/1 to: T28565 from: 1 to: 861
FT FT
FT FT      1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
FT FT      498 TACCGCATTCGACTACGCGCGGTGCGGATAT 533
FT FT
FT FT      seq_name: N_Geneseq_36:N60877
FT FT
FT FT      seq_documentation_block:
FT FT      ID N60877 standard; DNA; 1400 BP.
FT FT      AC N60877;
FT FT      DT 28-OCT-1991 (first entry)
FT FT      DE Sequence encoding Serratia phospholipase and promoter.
FT FT      KW Phospholipids; fatty acids; pN121; ss.
FT FT      OS Serratia sp.
FT FT      FH Key Location/Qualifiers
FT FT      FT promoter 1..215
FT FT      FT cds 216..1175
FT FT      FT /*tag= a
FT FT      FT /*tag= b
FT FT      FT /note= "Claim 23"
FT FT
FT FT      W08060743-A.
FT FT      PD 20-NOV-1986.
FT FT      PF 09-MAY-1986; DK0051.
FT FT      PR 10-MAY-1985; DK-002100.
FT FT      PR 23-DEC-1985; DK-006060.
FT FT      PR 09-MAY-1986; WO-DK0051.
FT FT      PR 09-JAN-1987; WO-000097.
FT FT      (BENA ) BENZON A AS.

```

```

Ratio: 4.000 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 58.333

alignment_block:
US-08-653-294-36 x T26331 ..

Align seg 1/1 to: T26331 from: 1 to: 382

1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
||||| ::| |||||
36 TACAGCGGTGAAC TACGCGCGCTGGCCCTCAGATAT 71

seq_name: N_Geneseq_36:Q99804

seq_documentation_block:
ID Q99804 standard; cDNA; 731 BP.
AC Q99804;
DT 20-JUN-1996 (first entry)
DE PR-1 like gene PR-1mz.
KW SAR; tobacco; protein-synthesis independent gene; cyclohexamide;
KW systemic acquired resistance response; anti-pathogen; plant protection;
KW maize; PR-1; ss.
OS Zea mays.
FH Key Location/Qualifiers
FT cds 40..531
FT CDS /*tag= a
W09519443-A2.
PD 20-JUL-1995.
PF 03-JAN-1995; IB0002.
PR 13-JAN-1994; US-181271.
PC (CIBA ) CIBA GEIGY AG.
PI Alexander DC, Ryals JA, Uknes SJ, Ward ER;
DR WPI: 95-263872/34.
DP P-PSDB; R91595.
PT New DNA contg. plant systemic acquired resistance genes - and
PT transgenic plants contg. them, impart disease and pest resistance,
PT also Arabidopsis gene promoter to control DNA transcription
PS Claim 21; Page 70; 85pp; English.
SC This sequence represents the DNA sequence of a maize PR-1 like gene,
CC PR-1mz. This sequence was isolated by screening a BPH-induced cDNA
CC library of maize. The library was screened using a probe matching to
CC PR-1 barley clone HVR1Bk. This sequence, Q99800-Q99803 and Q99805 are
CC all used in recombinant/chimaeric DNA molecules of the invention. The
CC sequences were isolated by differential screening of a cDNA library,
CC followed by analysis by Northern hybridisation to RNA in the presence
CC of cyclohexamide. The genes are used in the creation of
CC transgenic plants. All of these sequences confer anti-pathogenic
CC properties to transgenic plants. Transgenic expression of 2 or more o
CC the recombinant molecules of the invention that encode anti-pathogeni
CC proteins provides a synergistic increase in plant protection, and may
CC also offer protection against a wider range of pathogens.
SQ Sequence 731 BP; 186 A; 212 C; 199 G; 134 T;

alignment_scores:
Quality: 36.00 Length: 9
Ratio: 4.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 77.778

alignment_block:
US-08-653-294-36 x Q99804/rev ..

Align seg 1/1 to reverse of: Q99804 from: 1 to: 731

3 LeuAlaIleArgArgIleAlaLeuArg 11
||||| ::| |||||
231 CTGCGCGTGGCGCGCTAGCTGCGC 205

seq_name: N_Geneseq_36:Q44280

seq_documentation_block:
ID Q44280 standard; DNA; 1092 BP.
AC Q44280;

```

```

DT 09-DEC-1993 (first entry)
DE Pseudomonas cepacia DSM 3401 lipD gene.
KW Lipase; LipD; lipase modulator; limD; chaperone molecule;
KW lipolysis; detergent; ss.
OS Pseudomonas cepacia.
FFH Key Location/Qualifiers
FT 391
FT /tag= a
FT 480
FT /tag= b
FT 481
FT /tag= c
FT 712..714
FT /tag= d
FT 734
FT /tag= e
FT W09313200-A.
FT PN
FT 08-JUL-1993.
FT PD
FT 18-DEC-1992; DK0391.
FT PF
FT 20-DEC-1991; WO-DK0402.
FT PR
FT (NOVO) NOVO-NORDISK AS.
FT PA
FT Buckley CM, Diderichsen BK, Hobson A, Joergensen ST;
FT PI McConnell DJ;
FT WPI; 93-227318/28.
FT DR
FT P-PSDB; R39396.
FT DPT
FT denaturation and restructuring in presence of chaperone molecule
FT PS
FT Example 10; Page 43; 78pp; English.
FT CCC
FT Two genes were cloned and sequenced from Pseudomonas cepacia DSM
FT CCC 3401. The genes were designated lipD (Q44280) and limD (Q44281) and
FT CCC they code for a lipase and a lipase modulator protein, respectively.
FT CCC Due to the extreme GC content of the DNA, the sequence was difficult
FT CCC to determine (hence the "others" in the sequence). The limD start
FT CCC codon is positioned 3 bp downstream of the lipD stop codon. LipD and
FT CCC limD were found to be homologous to lipA and limA, respectively. In
FT CCC denaturation/renaturation experiments, limA chaperone protein was
FT CCC able to produce active LipD.
FT SQ Sequence 1092 BP; 188 A; 353 C; 378 G; 166 T;

alignment_scores:
Quality: 36.00 Length: 10
Ratio: 4.000 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-36 x Q44280/rev ..
Align seg 1/1 to reverse of: Q44280 from: 1 to: 1092

3 LeuAlaileArgArgileAlaLeuArgTyr 12
|||||:|||||
960 CTGGCGGTACAGCGCGCTGCACCTTCGATAC 931

seq_name: N_Geneseq_36:T17455

seq_documentation_block:
ID T17455 standard; cDNA; 24025 BP.
AC T17455;
DT 07-OCT-1996 (first entry)
DE Mutated BRCA1 genomic sequence from sample set MSKCC family 19921.
DE Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
DE antibody production; germline alteration; probe; lesion neoplasia; human;
DE gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FFH Key Location/Qualifiers
FT 256..355
FT /tag= a
FT /note= "exon 1"
FT 356..1512
FT /tag= b
FT /note= "intron 1"
FT misc_feature 1295

```



```

FT FT exon /note= "known polymorphic site"
FT FT 6572..6677
FT FT /tag= ab
FT FT /note= "exon 8"
FT FT 6678..9163
FT FT /tag= ac
FT FT /note= "intron 8"
FT FT 6823
FT FT /tag= ad
FT FT /note= "known polymorphic site"
FT FT 9106
FT FT /tag= ae
FT FT /note= "known polymorphic site"
FT FT 9164..9209
FT FT /tag= af
FT FT /note= "exon 9"
FT FT 9207
FT FT /tag= ag
FT FT /note= "known polymorphic site"
FT FT 9210..10530
FT FT /tag= ah
FT FT /note= "intron 9"
FT FT 9376
FT FT /tag= ai
FT FT /note= "known polymorphic site"
FT FT 10531..10607
FT FT /tag= aj
FT FT /note= "exon 10"
FT FT 10608..11397
FT FT /tag= ak
FT FT /note= "intron 10"
FT FT 11384..11396
FT FT /tag= al
FT FT /note= "indefinite interval within intron 10"
FT FT 11398..15023
FT FT /tag= am
FT FT /note= "exon 11"
FT FT 11908
FT FT /tag= an
FT FT /note= "known polymorphic site"
FT FT 11994
FT FT /tag= ao
FT FT /note= "known polymorphic site"
FT FT 12952
FT FT /tag= ap
FT FT /note= "known polymorphic site"
FT FT 13004
FT FT /tag= aq
FT FT /note= "known polymorphic site"
FT FT 13009
FT FT /tag= ar
FT FT /note= "known polymorphic site"
FT FT 13048
FT FT /tag= as
FT FT /note= "known polymorphic site"
FT FT 13238
FT FT /tag= at
FT FT /note= "known polymorphic site"
FT FT 13448
FT FT /tag= au
FT FT /note= "known polymorphic site"
FT FT 13539
FT FT /tag= av
FT FT /note= "known polymorphic site"
FT FT 13951
FT FT /tag= aw
FT FT /note= "known polymorphic site"
FT FT 14041
FT FT /tag= ax
FT FT /note= "known polymorphic site"
FT FT 14046
FT FT /tag= ay
FT FT /note= "known polymorphic site"

```

```

FT FT misc_feature 14475
FT FT /tag= az
FT FT /note= "known polymorphic site"
FT FT 14874
FT FT /tag= ba
FT FT /note= "known polymorphic site"
FT FT 14891
FT FT /tag= bb
FT FT /note= "known polymorphic site"
FT FT 14966
FT FT /tag= bc
FT FT /note= "known polymorphic site"
FT FT 15024..15424
FT FT /tag= bd
FT FT /note= "intron 11"
FT FT 15284
FT FT /tag= be
FT FT /note= "known polymorphic site"
FT FT 15425..15511
FT FT /tag= bf
FT FT /note= "exon 12"
FT FT 15512..15952
FT FT /tag= bg
FT FT /note= "intron 12"
FT FT 15647..15659
FT FT /tag= bh
FT FT /note= "indefinite interval within intron 12"
FT FT 15953..16126
FT FT /tag= bi
FT FT /note= "exon 13"
FT FT 16077
FT FT /tag= bj
FT FT /note= "known polymorphic site"
FT FT 16127..16565
FT FT /tag= bk
FT FT /note= "intron 13"
FT FT 16243
FT FT /tag= bl

alignment_scores:
Quality: 36.00 Length: 11
Ratio: 3.600 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
US-08-653-294-36 x TI7455/rev ...
Align seg 1/1 to reverse of: TI7455 from: 1 to: 24025

2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
11313 AGATTGCCCATCAGAAACTGGTCTCTGATGATC 11281

seq_name: N_Geneseq_36:TI7515
seq_documentation_block:
ID TI7515 standard; cDNA: 24025 BP.
AC TI7515;
DE Mutated BRCA1 genomic sequence from PM15.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens. Location/Qualifiers
FH Key 256..355
FT exon /tag= a
FT FT /note= "exon 1"
FT FT 356..1512
FT FT /tag= b
FT FT /note= "intron 1"
FT FT 1295
FT FT /tag= c

```

```
FT FT /note= "known polymorphic site"
FT FT 1513. .1611
FT FT /*tag= d
FT FT /note= "exon 2"
FT FT 1612. .2206
FT FT /*tag= e
FT FT /note= "intron 2"
FT FT 1925. .1937
FT FT /*tag= f
FT FT /note= "indefinite interval within intron 2"
FT FT 2141
FT FT /*tag= g
FT FT /note= "known polymorphic site"
FT FT 2207. .2260
FT FT /*tag= h
FT FT /note= "exon 3"
FT FT 2261. .2677
FT FT /*tag= i
FT FT /note= "intron 3"
FT FT 2569. .2581
FT FT /*tag= j
FT FT /note= "indefinite interval within intron 3"
FT FT 2678. .2788
FT FT /*tag= k
FT FT /note= "exon 4"
FT FT 2725
FT FT /*tag= l
FT FT /note= "known polymorphic site"
FT FT 2789. .3328
FT FT /*tag= m
FT FT /note= "intron 4"
FT FT 3063. .3075
FT FT /*tag= n
FT FT /note= "indefinite interval within intron 4"
FT FT 3329. .3406
FT FT /*tag= o
FT FT /note= "exon 5"
FT FT 3407. .3813
FT FT /*tag= p
FT FT /note= "intron 5"
FT FT 3598. .3610
FT FT /*tag= q
FT FT /note= "indefinite interval within intron 5"
FT FT 3653
FT FT /*tag= r
FT FT /note= "known polymorphic site"
FT FT 3814. .3902
FT FT /*tag= s
FT FT /note= "exon 6"
FT FT 3903. .4224
FT FT /*tag= t
FT FT /note= "intron 6"
FT FT 4076. .4088
FT FT /*tag= u
FT FT /note= "indefinite interval within intron 6"
FT FT 4225. .4364
FT FT /*tag= v
FT FT /note= "exon 7"
FT FT 4365. .6571
FT FT /*tag= w
FT FT /note= "intron 7"
FT FT 4391. .4392
FT FT /*tag= x
FT FT /note= "known polymorphic site"
FT FT 4602. .4614
FT FT /*tag= y
FT FT /note= "indefinite interval within intron 7"
FT FT 6538
FT FT /*tag= z
FT FT /note= "known polymorphic site"
FT FT 6572. .6677
FT FT /*tag= aa
FT FT /note= "exon 8"
FT FT

FT FT intron
FT FT 6678. .9163
FT FT /*tag= ab
FT FT /note= "intron 8"
FT FT 6823
FT FT /*tag= ac
FT FT /note= "known polymorphic site"
FT FT 9106
FT FT /*tag= ad
FT FT /note= "site of 1 nucleotide deletion at known
FT FT polymorphic site"
FT FT 9163. .9208
FT FT /*tag= ae
FT FT /note= "exon 9"
FT FT 9206
FT FT /*tag= af
FT FT /note= "known polymorphic site"
FT FT 9209. .10529
FT FT /*tag= ag
FT FT /note= "intron 9"
FT FT 9375
FT FT /*tag= ah
FT FT /note= "known polymorphic site"
FT FT 10530. .10606
FT FT /*tag= ai
FT FT /note= "exon 10"
FT FT 10607. .11596
FT FT /*tag= aj
FT FT /note= "intron 10"
FT FT 11383. .11395
FT FT /*tag= ak
FT FT /note= "indefinite interval within intron 10"
FT FT 11597. .15022
FT FT /*tag= al
FT FT /note= "exon 11"
FT FT 11907
FT FT /*tag= am
FT FT /note= "known polymorphic site"
FT FT 11993
FT FT /*tag= an
FT FT /note= "known polymorphic site"
FT FT 12951
FT FT /*tag= ao
FT FT /note= "known polymorphic site"
FT FT 13003
FT FT /*tag= ap
FT FT /note= "known polymorphic site"
FT FT 13008
FT FT /*tag= aq
FT FT /note= "known polymorphic site"
FT FT 13047
FT FT /*tag= ar
FT FT /note= "known polymorphic site"
FT FT 13237
FT FT /*tag= as
FT FT /note= "known polymorphic site"
FT FT 13447
FT FT /*tag= at
FT FT /note= "known polymorphic site"
FT FT 13538
FT FT /*tag= au
FT FT /note= "known polymorphic site"
FT FT 13950
FT FT /*tag= av
FT FT /note= "known polymorphic site"
FT FT 14040
FT FT /*tag= aw
FT FT /note= "known polymorphic site"
FT FT 14045
FT FT /*tag= ax
FT FT /note= "known polymorphic site"
FT FT 14474
FT FT /*tag= ay
FT FT /note= "known polymorphic site"
FT FT
```

```

FT misc_feature 14873 /tag= az
FT /note= "known polymorphic site"
FT 14890 /tag= ba
FT /note= "known polymorphic site"
FT 14965 /tag= bb
FT /note= "known polymorphic site"
FT 15023 /tag= bc
FT /note= "intron 11"
FT 15283 /tag= bd
FT /note= "known polymorphic site"
FT 15424 /tag= be
FT /note= "exon 12"
FT 15511 /tag= bf
FT /note= "intron 12"
FT 15646 /tag= bg
FT /note= "indefinite interval within intron 12"
FT 15952 /tag= bh
FT /note= "exon 13"
FT 16076 /tag= bi
FT /note= "known polymorphic site"
FT 16126 /tag= bj
FT /note= "intron 13"
FT 16242 /tag= bk
FT /note= "known polymorphic site"
FT 16369 /tag= bl
FT /note= "known polymorphic site"

alignment_scores:
  Quality: 36.00 Length: 11
  Ratio: 3.600 Gaps: 0
  Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
  US-08-653-294-36 x T17515/rev ..
  Align seg 1/1 to reverse of: T17515 from: 1 to: 24025

      2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
      |||||
11313 AGATTGGCCATCAGAAACTGGTCTGATGTAC 11281

seq_name: N_Geneseq_36.T18325

seq_documentation_block:
ID T18325 standard; DNA; 24026 BP.
AC T18325;
DE 05-JUN-1996 (first entry)
DE BRCA1, human breast and ovarian cancer predisposing gene.
KW BRCA1; breast cancer; ovarian cancer; predisposing gene;
KW susceptibility gene; diagnosis; prognosis; gene therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT intron 1..55
FT /tag= a
FT exon 56..155
FT /tag= b
FT intron 156..1512
FT /tag= c
FT exon 1513..1611
FT /tag= d
FT intron 1612..2206
FT /tag= e

```

```

FT exon /note= "n at 1925-1937 represent an indefinite
FT interval within the intron"
FT /tag= f
FT 2261..2677
FT /tag= g
FT /note= "n at 2569-2581 represent an indefinite
FT interval within the intron"
FT 2678..2788
FT /tag= h
FT 2789..3328
FT /tag= i
FT /note= "n at 3063-3075 represent an indefinite
FT interval within the intron"
FT 3329..3406
FT /tag= j
FT 3407..3813
FT /tag= k
FT /note= "n at 3598-3610 represent an indefinite
FT interval within the intron"
FT 3814..3902
FT /tag= l
FT 3903..4224
FT /tag= m
FT /note= "n at 4076-4088 represent an indefinite
FT interval within the intron"
FT 4225..4364
FT /tag= n
FT 4365..6571
FT /tag= o
FT /note= "n at 4602-4614 represent an indefinite
FT interval within the intron"
FT 6572..6677
FT /tag= p
FT 6678..9163
FT /tag= q
FT 9164..9207
FT /tag= r
FT 9208..10530
FT /tag= s
FT 10531..10607
FT /tag= t
FT 10608..11597
FT /tag= u
FT /note= "n at 11383-11396 represent an indefinite
FT interval within the intron"
FT 11598..15023
FT /tag= v
FT 15024..15424
FT /tag= w
FT 15425..15511
FT /tag= x
FT 15512..15952
FT /tag= y
FT /note= "n at 15647-15659 represent an indefinite
FT interval within the intron"
FT 15953..16126
FT /tag= z
FT 16127..16565
FT /tag= aa
FT /note= "n at 16370-16382 represent an indefinite
FT interval within the intron"
FT 16566..16692
FT /tag= ab
FT 16693..17535
FT /tag= ac
FT /note= "n at 17290-17302 represent an indefinite
FT interval within the intron"
FT 17536..17726
FT /tag= ad
FT 17727..18416
FT /tag= ae
FT /note= "n at 18299-18312 represent an indefinite

```

```

FT exon          interval within the intron"
FT 18417..18787  /*tag= af
FT 18788..19298  /*tag= ag
FT /note= "n at 18952-18964 represent an indefinite
FT 19299..19386  interval within the intron"
FT /tag= ah
FT 19387..20190  /*tag= ai
FT /note= "n at 19887-19899 represent an indefinite
FT 20191..20267  interval within the intron"
FT /tag= aj
FT 20268..21094  /*tag= ak
FT /note= "n at 20767-20779 represent an indefinite
FT 21095..21135  interval within the intron"
FT /tag= al
FT 21136..21583  /*tag= am
FT /note= "n at 21341-21353 represent an indefinite
FT 21584..21667  interval within the intron"
FT /tag= an
FT 21668..22233  /*tag= ao
FT /note= "n at 21921-21933 represent an indefinite
FT 22234..22288  interval within the intron"
FT /tag= ap
FT 22289..22832  /*tag= aq
FT /note= "n at 22567-22579 represent an indefinite
FT 22833..22906  interval within the intron"
FT /tag= ar
FT 22907..23287  /*tag= as
FT /note= "n at 23050-23062 represent an indefinite
FT 23288..23348  interval within the intron"
FT /tag= at
FT 23349..23698  /*tag= au
FT /note= "n at 23580-23592 represent an indefinite
FT 23699..24026  interval within the intron"
FT /tag= av
FT misc_feature 2725  /*tag= aw
FT /note= "polymorphic site"
FT 3653  /*tag= ax
FT /note= "polymorphic site"
FT 4391  /*tag= ay
FT /note= "polymorphic site"
FT 4392  /*tag= az
FT /note= "polymorphic site"
FT 6823  /*tag= ba
FT /note= "polymorphic site"
FT 9106  /*tag= bb
FT /note= "polymorphic site"
FT 9207  /*tag= bc
FT /note= "polymorphic site"
FT 9376

```

```

FT /tag= bd
FT /note= "polymorphic site"
FT 11908  /*tag= be
FT /note= "polymorphic site"
FT 11994  /*tag= bf
FT /note= "polymorphic site"
FT 12952  /*tag= bg
FT /note= "polymorphic site"
FT 13004  /*tag= bh
FT /note= "polymorphic site"
FT 13009  /*tag= bi
FT /note= "polymorphic site"
FT 13048  /*tag= bj
FT /note= "polymorphic site"
FT 13238  /*tag= bk
FT /note= "polymorphic site"
FT 13448  /*tag= bl
FT /note= "polymorphic site"
FT 13539  /*tag= bm
FT /note= "polymorphic site"
FT 13951  /*tag= bn
FT /note= "polymorphic site"
FT 14041  /*tag= bo
FT /note= "polymorphic site"
FT 14046  /*tag= bo
FT /note= "polymorphic site"

alignment_scores:
Quality: 36.00 Length: 11
Ratio: 3.600 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
US-08-653-294-36 x T18325/rev ...
Align seg 1/1 to reverse of: T18325 from: 1 to: 24026

2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
11314 AGATTGGCATCAGAAACTGGTCTGTATGTAC 11282

seq_name: N_Geneseq_36:T17512

seq_documentation_block:
ID T17512 standard; cDNA; 24026 BP.
AC T17512;
DT 04-OCT-1996 (first entry)
DE Mutated BRCA1 genomic sequence from PM04.
KW Cancer therapy; breast and ovarian cancer
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 256..355
FT /tag= a
FT /note= "exon 1"
FT 356..1512
FT /tag= b
FT /note= "intron 1"
FT 1295
FT /tag= c
FT /note= "known polymorphic site"
FT 1513..1611

```

```
FT FT /*tag= d /note= "exon 2" /note= "intron 8"
FT FT 1612..2206 /*tag= e /note= "intron 2" /*tag= ac
FT FT /*tag= f /note= "indefinite interval within intron 2" /*tag= "known polymorphic site"
FT FT 1925..1937 /*tag= g /note= "known polymorphic site"
FT FT 2141 /*tag= h /note= "exon 3"
FT FT /*tag= i /note= "intron 3"
FT FT 2569..2581 /*tag= j /note= "indefinite interval within intron 3"
FT FT /*tag= k /note= "exon 4"
FT FT 2725 /*tag= l /note= "known polymorphic site"
FT FT 2789..3328 /*tag= m /note= "intron 4"
FT FT /*tag= n /note= "indefinite interval within intron 4"
FT FT 3063..3075 /*tag= o /note= "exon 5"
FT FT 3329..3406 /*tag= p /note= "intron 5"
FT FT 3407..3813 /*tag= q /note= "indefinite interval within intron 5"
FT FT 3598..3610 /*tag= r /note= "known polymorphic site"
FT FT 3653 /*tag= s /note= "known polymorphic site"
FT FT 3814..3902 /*tag= t /note= "intron 6"
FT FT /*tag= u /note= "indefinite interval within intron 6"
FT FT 3903..4224 /*tag= v /note= "exon 7"
FT FT 4076..4088 /*tag= w /note= "intron 7"
FT FT 4225..4364 /*tag= x /note= "known polymorphic site"
FT FT /*tag= y /note= "indefinite interval within intron 7"
FT FT 4365..4614 /*tag= z /note= "known polymorphic site"
FT FT 4602..4614 /*tag= aa /note= "exon 8"
FT FT /*tag= ab /note= "known polymorphic site"
FT FT 6538 /*tag= ac /note= "intron 10"
FT FT 6572..6677 /*tag= ad /note= "indefinite interval within intron 10"
FT FT /*tag= ae /note= "exon 11"
FT FT 6678..9163 /*tag= af /note= "known polymorphic site"
FT FT /*tag= ag /note= "intron 10"
FT FT 9164..9209 /*tag= ah /note= "known polymorphic site"
FT FT /*tag= ai /note= "exon 10"
FT FT 9207 /*tag= aj /note= "intron 10"
FT FT /*tag= ak /note= "known polymorphic site"
FT FT 9210..10530 /*tag= al /note= "exon 11"
FT FT /*tag= am /note= "known polymorphic site"
FT FT 9376 /*tag= an /note= "known polymorphic site"
FT FT 10531..10607 /*tag= ao /note= "known polymorphic site"
FT FT /*tag= ap /note= "known polymorphic site"
FT FT 10608..11597 /*tag= aq /note= "known polymorphic site"
FT FT /*tag= ar /note= "known polymorphic site"
FT FT 11384..11396 /*tag= as /note= "known polymorphic site"
FT FT /*tag= at /note= "known polymorphic site"
FT FT 11598..15023 /*tag= au /note= "known polymorphic site"
FT FT /*tag= av /note= "known polymorphic site"
FT FT 11908 /*tag= aw /note= "known polymorphic site"
FT FT 11994 /*tag= ax /note= "known polymorphic site"
FT FT 12952 /*tag= ay /note= "known polymorphic site"
FT FT /*tag= az /note= "known polymorphic site"
FT FT 13004 /*tag= ba /note= "known polymorphic site"
FT FT /*tag= bb /note= "known polymorphic site"
FT FT 13009 /*tag= bc /note= "known polymorphic site"
FT FT 13048 /*tag= bd /note= "known polymorphic site"
FT FT 13238 /*tag= be /note= "known polymorphic site"
FT FT /*tag= bf /note= "known polymorphic site"
FT FT 13448 /*tag= bg /note= "known polymorphic site"
FT FT /*tag= bh /note= "known polymorphic site"
FT FT 13539 /*tag= bi /note= "known polymorphic site"
FT FT /*tag= bj /note= "known polymorphic site"
FT FT 13951 /*tag= bk /note= "known polymorphic site"
FT FT /*tag= bl /note= "known polymorphic site"
FT FT 14041 /*tag= bm /note= "known polymorphic site"
FT FT /*tag= bn /note= "known polymorphic site"
FT FT 14046 /*tag= bo /note= "known polymorphic site"
FT FT /*tag= bp /note= "known polymorphic site"
FT FT 14475 /*tag= bq /note= "known polymorphic site"
FT FT /*tag= br /note= "known polymorphic site"
FT FT 14874 /*tag= bs /note= "known polymorphic site"
FT FT /*tag= bt /note= "known polymorphic site"
```

```

FT misc_feature 14891 /*tag= ba /*tag= "known polymorphic site"
FT misc_feature 14966 /*tag= bb /*tag= "known polymorphic site"
FT intron 15024 /*tag= bc /*tag= "known polymorphic site"
FT mutation 15284 /*tag= bd /*tag= "known polymorphic site"
FT exon 15425 /*tag= be /*tag= "known polymorphic site"
FT intron 15512 /*tag= bf /*tag= "known polymorphic site"
FT misc_feature 15647 /*tag= bg /*tag= "known polymorphic site"
FT exon 15953 /*tag= bh /*tag= "known polymorphic site"
FT misc_feature 16077 /*tag= bi /*tag= "known polymorphic site"
FT intron 16127 /*tag= bj /*tag= "known polymorphic site"
FT misc_feature 16243 /*tag= bk /*tag= "known polymorphic site"
FT misc_feature 16370 /*tag= bl /*tag= "known polymorphic site"

alignment_scores:
  Quality: 36.00 Length: 11
  Ratio: 3.600 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
US-08-653-294-36 x T17512/rev ..
Align seg 1/1 to reverse of: T17512 from: 1 to: 24026

2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
11314 AGATTGGCCATCAGAAACTGGTTCGTATGATC 11282

seq_name: N_Geneseq_36:T17513
seq_documentation_block:
ID T17513 standard; cDNA; 24026 BP.
DE Mutated BRCA1 genomic sequence from PM05.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 256..355
FT /*tag= a
FT /*tag= "exon 1"
FT intron 356..1512
FT /*tag= b
FT /*tag= "intron 1"
FT misc_feature 1295
FT /*tag= c
FT /*tag= "known polymorphic site"
FT exon 1513..1611
FT /*tag= d

```

```

FT intron 1612..2206
FT /*tag= e
FT /*tag= "intron 2"
FT misc_feature 1925..1937
FT /*tag= f
FT /*tag= "indefinite interval within intron 2"
FT misc_feature 2141
FT /*tag= g
FT /*tag= "known polymorphic site"
FT exon 2207..2260
FT /*tag= h
FT /*tag= "exon 3"
FT intron 2261..2677
FT /*tag= i
FT /*tag= "intron 3"
FT misc_feature 2569..2581
FT /*tag= j
FT /*tag= "indefinite interval within intron 3"
FT exon 2678..2788
FT /*tag= k
FT /*tag= "exon 4"
FT misc_feature 2725
FT /*tag= l
FT /*tag= "known polymorphic site"
FT intron 2789..3328
FT /*tag= m
FT /*tag= "intron 4"
FT misc_feature 3063..3075
FT /*tag= n
FT /*tag= "indefinite interval within intron 4"
FT exon 3329..3406
FT /*tag= o
FT /*tag= "exon 5"
FT intron 3407..3813
FT /*tag= p
FT /*tag= "intron 5"
FT misc_feature 3598..3610
FT /*tag= q
FT /*tag= "indefinite interval within intron 5"
FT misc_feature 3653
FT /*tag= r
FT /*tag= "known polymorphic site"
FT exon 3814..3902
FT /*tag= s
FT /*tag= "exon 6"
FT intron 3903..4224
FT /*tag= t
FT /*tag= "intron 6"
FT misc_feature 4076..4088
FT /*tag= u
FT /*tag= "indefinite interval within intron 6"
FT exon 4225..4364
FT /*tag= v
FT /*tag= "exon 7"
FT intron 4365..46571
FT /*tag= w
FT /*tag= "intron 7"
FT misc_feature 4391..4392
FT /*tag= x
FT /*tag= "known polymorphic site"
FT misc_feature 4602..4614
FT /*tag= y
FT /*tag= "indefinite interval within intron 7"
FT misc_feature 6538
FT /*tag= z
FT /*tag= "known polymorphic site"
FT exon 6572..6677
FT /*tag= aa
FT /*tag= "exon 8"
FT intron 6678..9163
FT /*tag= ab
FT /*tag= "intron 8"

```

```
FT misc_feature 6823 /*tag= ac /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 9106 /*tag= ad /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 9164.9209 /*tag= ae /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 9207 /*tag= af /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 9210.10530 /*tag= ag /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 9376 /*tag= ah /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 10531.10607 /*tag= ai /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 10608.11597 /*tag= aj /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 11384.11396 /*tag= ak /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 11598.15023 /*tag= al /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 11908 /*tag= am /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 11994 /*tag= an /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 12952 /*tag= ao /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13004 /*tag= ap /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13009 /*tag= aq /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13048 /*tag= ar /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13238 /*tag= as /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13448 /*tag= at /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13539 /*tag= au /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 13951 /*tag= av /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 14041 /*tag= aw /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 14046 /*tag= ax /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 14475 /*tag= ay /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 14874 /*tag= az /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 14891
```

```
FT /*tag= ba /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 14966 /*tag= bb /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 15024.15424 /*tag= bc /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 15284 /*tag= bd /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 15425.15511 /*tag= be /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 15512.15952 /*tag= bf /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 15647.15659 /*tag= bg /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 15953.16126 /*tag= bh /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 16077 /*tag= bi /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 16127.16565 /*tag= bj /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 16243 /*tag= bk /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT 16370.16382 /*tag= bl /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"

alignment_scores:
  Quality: 36.00 Length: 11
  Ratio: 3.600 Gaps: 0
  Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
  US-08-653-294-36 x T17513/rev ..
  Align seg 1/1 to reverse of: T17513 from: 1 to: 24026
  2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
  |||||
  11314 AGATTGCCCATCAGAAACTGGTCTGTGATGATC 11282

seq_name: N_Geneseq_36:T17514

seq_documentation_block:
  ID T17514 standard; cDNA; 24026 BP.
  AC T17514;
  DT 04-OCT-1996 (first entry)
  DE Mutated BRCAL genomic sequence from PM11.
  KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
  KW antibody production; germline alteration; protein replacement therapy; protein mimetic; BRCAL; ds.
  OS Homo sapiens.
  FH Key Location/Qualifiers
  FT exon 256..355
  FT /*tag= a /*tag= "exon 1"
  FT 356..1512 /*tag= b /*tag= "intron 1"
  FT 1295 /*tag= c /*tag= "known polymorphic site"
  FT /*tag= d /*tag= "exon 2"
  FT 1513..1611
```

```
FT intron 1612..2206 /tag= e /note= "intron 2" /tag= ac
FT misc_feature 1925..1937 /tag= f /note= "known polymorphic site"
FT misc_feature 2141 /tag= g /note= "indefinite interval within intron 2"
FT exon 2207..2260 /tag= h /note= "known polymorphic site"
FT intron 2261..2677 /tag= i /note= "exon 3"
FT misc_feature 2369..2581 /tag= j /note= "intron 3"
FT exon 2678..2788 /tag= k /note= "indefinite interval within intron 3"
FT misc_feature 2725 /tag= l /note= "exon 4"
FT intron 2789..3328 /tag= m /note= "known polymorphic site"
FT misc_feature 3063..3075 /tag= n /note= "intron 4"
FT exon 3329..3406 /tag= o /note= "indefinite interval within intron 4"
FT intron 3407..3813 /tag= p /note= "exon 5"
FT misc_feature 3598..3610 /tag= q /note= "intron 5"
FT misc_feature 3653 /tag= r /note= "indefinite interval within intron 5"
FT exon 3814..3902 /tag= s /note= "known polymorphic site"
FT intron 3903..4224 /tag= t /note= "exon 6"
FT misc_feature 4076..4088 /tag= u /note= "intron 6"
FT exon 4225..4364 /tag= v /note= "indefinite interval within intron 6"
FT intron 4365..6571 /tag= w /note= "exon 7"
FT misc_feature 4391..4392 /tag= x /note= "intron 7"
FT misc_feature 4602..4614 /tag= y /note= "known polymorphic site"
FT misc_feature 6538 /tag= z /note= "indefinite interval within intron 7"
FT exon 6572..6677 /tag= aa /note= "known polymorphic site"
FT intron 6678..9163 /tag= ab /note= "exon 8"
FT misc_feature 6823 /tag= ac /note= "intron 8"
FT /tag= ad /note= "known polymorphic site"
FT /tag= ae /note= "exon 9"
FT /tag= af /note= "known polymorphic site"
FT /tag= ag /note= "intron 9"
FT /tag= ah /note= "known polymorphic site"
FT /tag= ai /note= "exon 10"
FT /tag= aj /note= "intron 10"
FT /tag= ak /note= "indefinite interval within intron 10"
FT /tag= al /note= "exon 11"
FT /tag= am /note= "known polymorphic site"
FT /tag= an /note= "known polymorphic site"
FT /tag= ao /note= "known polymorphic site"
FT /tag= ap /note= "known polymorphic site"
FT /tag= aq /note= "known polymorphic site"
FT /tag= ar /note= "known polymorphic site"
FT /tag= as /note= "known polymorphic site"
FT /tag= at /note= "known polymorphic site"
FT /tag= au /note= "known polymorphic site"
FT /tag= av /note= "known polymorphic site"
FT /tag= aw /note= "known polymorphic site"
FT /tag= ax /note= "known polymorphic site"
FT /tag= ay /note= "known polymorphic site"
FT /tag= az /note= "known polymorphic site"
FT /tag= ba
```



```

FT      misc_feature      /note= "known polymorphic site"
FT      14966
FT      /tag= bb
FT      /note= "known polymorphic site"
FT      15024..15424
FT      /tag= bc
FT      /note= "intron 11"
FT      15284
FT      /tag= bd
FT      /note= "known polymorphic site"
FT      15425..15511
FT      /tag= be
FT      /note= "exon 12"
FT      15512..15952
FT      /tag= bf
FT      /note= "intron 12"
FT      15847..15659
FT      /tag= bg
FT      /note= "indefinite interval within intron 12"
FT      15953..16126
FT      /tag= bh
FT      /note= "exon 13"
FT      16077
FT      /tag= bi
FT      /note= "known polymorphic site"
FT      16127..16565
FT      /tag= bj
FT      /note= "intron 13"
FT      16243
FT      /tag= bk
FT      /note= "known polymorphic site"
FT      16370..16382
FT      /tag= bl

alignment_scores:
  Quality: 36.00      Length: 11
  Ratio: 3.600      Gaps: 0
  Percent Similarity: 90.909      Percent Identity: 63.636

alignment_block:
  US-08-653-294-36 x T17514/rev ..

  Align seg 1/1 to reverse of: T17514 from: 1 to: 24026

      2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
      |||||
      11314 AGATTGGCATCAGAAACTGTTCTGATGATAC 11282

seq_name: N_Geneseq_36:T17516

seq_documentation_block:
  ID T17516 standard; cDNA; 24026 BP.
  AC T17516;
  DT 04-OCT-1996 (first entry)
  DE Mutated BRCA1 genomic sequence from PM16.
  KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
  KW antibody production; germline alteration; probe; lesion neoplasia; human;
  KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
  OS Homo sapiens.
  FH Key      Location/Qualifiers
  FT      256..355
  FT      /tag= a
  FT      /note= "exon 1"
  FT      356..1512
  FT      /tag= b
  FT      /note= "intron 1"
  FT      1295
  FT      /tag= c
  FT      /note= "known polymorphic site"
  FT      1513..1611
  FT      /tag= d
  FT      /note= "exon 2"
  FT      1612..2206

```

```

FT      /tag= e
FT      /note= "intron 2"
FT      1925..1937
FT      /tag= f
FT      /note= "indefinite interval within intron 2"
FT      2141
FT      /tag= g
FT      /note= "known polymorphic site"
FT      2207..2260
FT      /tag= h
FT      /note= "exon 3"
FT      2261..2677
FT      /tag= i
FT      /note= "intron 3"
FT      2569..2581
FT      /tag= j
FT      /note= "indefinite interval within intron 3"
FT      2678..2788
FT      /tag= k
FT      /note= "exon 4"
FT      2725
FT      /tag= l
FT      /note= "known polymorphic site"
FT      2789..3328
FT      /tag= m
FT      /note= "intron 4"
FT      3063..3075
FT      /tag= n
FT      /note= "indefinite interval within intron 4"
FT      3329..3406
FT      /tag= o
FT      /note= "exon 5"
FT      3407..3813
FT      /tag= p
FT      /note= "intron 5"
FT      3598..3610
FT      /tag= q
FT      /note= "indefinite interval within intron 5"
FT      3653
FT      /tag= r
FT      /note= "known polymorphic site"
FT      3814..3902
FT      /tag= s
FT      /note= "exon 6"
FT      3903..4224
FT      /tag= t
FT      /note= "intron 6"
FT      4076..4088
FT      /tag= u
FT      /note= "indefinite interval within intron 6"
FT      4225..4364
FT      /tag= v
FT      /note= "exon 7"
FT      4365..6571
FT      /tag= w
FT      /note= "intron 7"
FT      4391..4392
FT      /tag= x
FT      /note= "known polymorphic site"
FT      4602..4614
FT      /tag= y
FT      /note= "indefinite interval within intron 7"
FT      6538
FT      /tag= z
FT      /note= "known polymorphic site"
FT      6572..6677
FT      /tag= aa
FT      /note= "exon 8"
FT      6678..9163
FT      /tag= ab
FT      /note= "intron 8"
FT      6823
FT      /tag= ac

```

```
FT misc_feature /note= "known polymorphic site"
FT 9106 /tag= ad
FT /note= "known polymorphic site"
FT 9164 .9209 /tag= ae
FT /note= "exon 9"
FT 9207 /tag= af
FT /note= "known polymorphic site"
FT 9210 .10530 /tag= ag
FT /note= "intron 9"
FT 9376 /tag= ah
FT /note= "known polymorphic site"
FT 10531 .10607 /tag= ai
FT /note= "exon 10"
FT 10608 .11597 /tag= aj
FT /note= "intron 10"
FT 11384 .11396 /tag= ak
FT /note= "indefinite interval within intron 10"
FT 11598 .15023 /tag= al
FT /note= "exon 11"
FT 11908 /tag= am
FT /note= "known polymorphic site"
FT 11994 /tag= an
FT /note= "known polymorphic site"
FT 12952 /tag= ao
FT /note= "known polymorphic site"
FT 13004 /tag= ap
FT /note= "known polymorphic site"
FT 13009 /tag= aq
FT /note= "known polymorphic site"
FT 13048 /tag= ar
FT /note= "known polymorphic site"
FT 13238 /tag= as
FT /note= "known polymorphic site"
FT 13448 /tag= at
FT /note= "known polymorphic site"
FT 13539 /tag= au
FT /note= "known polymorphic site"
FT 13951 /tag= av
FT /note= "known polymorphic site"
FT 14041 /tag= aw
FT /note= "known polymorphic site"
FT 14046 /tag= ax
FT /note= "known polymorphic site"
FT 14475 /tag= ay
FT /note= "known polymorphic site"
FT 14874 /tag= az
FT /note= "known polymorphic site"
FT 14891 /tag= ba
FT /note= "known polymorphic site"
```

```
FT misc_feature 14966 /tag= bb
FT /note= "known polymorphic site"
FT 15024 .15424 /tag= bc
FT /note= "intron 11"
FT 15284 /tag= bd
FT /note= "known polymorphic site"
FT 15425 .15511 /tag= be
FT /note= "exon 12"
FT 15512 .15952 /tag= bf
FT /note= "intron 12"
FT 15847 .15659 /tag= bg
FT /note= "indefinite interval within intron 12"
FT 15953 .16126 /tag= bh
FT /note= "exon 13"
FT 16077 /tag= bi
FT /note= "known polymorphic site"
FT 16127 .16565 /tag= bj
FT /note= "intron 13"
FT 16243 /tag= bk
FT /note= "known polymorphic site"
FT 16370 .16382 /tag= bl
FT /note= "known polymorphic site"

alignment_scores:
  Quality: 36.00 Length: 11
  Ratio: 3.600 Gaps: 0
  Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
US-08-653-294-36 x T17516/rev ..

Align seg 1/1 to reverse of: T17516 from: 1 to: 24026

      2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
      |||||
11314 AGATTGGCCATCAGAAACTGGTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17517

seq_documentation_block:
ID T17517 standard; cDNA; 24026 BP.
AC T17517;
DT 04-OCR-1996 (first entry)
DE Mutated BRCA1 genomic sequence from PMA02.1.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 256..355
FT /tag= a
FT /note= "exon 1"
FT 356..1512 /tag= b
FT /note= "intron 1"
FT 1295 /tag= c
FT /note= "G to A mutation at known polymorphic site"
FT 1513..1611 /tag= d
FT /note= "exon 2"
FT 1612..2206 /tag= e
FT
```

```
FT FT misc_feature /note="intron 2"
FT FT 1925.1937
FT FT /tag=f
FT FT /note="indefinite interval within intron 2"
FT FT 2141
FT FT /tag=g
FT FT /note="known polymorphic site"
FT FT 2207.2260
FT FT /tag=h
FT FT /note="exon 3"
FT FT 2281.2677
FT FT /tag=i
FT FT /note="intron 3"
FT FT 2569.2581
FT FT /tag=j
FT FT /note="indefinite interval within intron 3"
FT FT 2678.2798
FT FT /tag=k
FT FT /note="exon 4"
FT FT 2725
FT FT /tag=l
FT FT /note="known polymorphic site"
FT FT 2789.3328
FT FT /tag=m
FT FT /note="intron 4"
FT FT 3063.3075
FT FT /tag=n
FT FT /note="indefinite interval within intron 4"
FT FT 3329.3406
FT FT /tag=o
FT FT /note="exon 5"
FT FT 3407.3813
FT FT /tag=p
FT FT /note="intron 5"
FT FT 3598.3610
FT FT /tag=q
FT FT /note="indefinite interval within intron 5"
FT FT 3653
FT FT /tag=r
FT FT /note="known polymorphic site"
FT FT 3814.3902
FT FT /tag=s
FT FT /note="exon 6"
FT FT 3903.4224
FT FT /tag=t
FT FT /note="intron 6"
FT FT 4076.4088
FT FT /tag=u
FT FT /note="indefinite interval within intron 6"
FT FT 4225.4364
FT FT /tag=v
FT FT /note="exon 7"
FT FT 4385.6571
FT FT /tag=w
FT FT /note="intron 7"
FT FT 4391.4392
FT FT /tag=x
FT FT /note="known polymorphic site"
FT FT 4602.4614
FT FT /tag=y
FT FT /note="indefinite interval within intron 7"
FT FT 6538
FT FT /tag=z
FT FT /note="known polymorphic site"
FT FT 6572.6677
FT FT /tag=aa
FT FT /note="exon 8"
FT FT 6678.9163
FT FT /tag=ab
FT FT /note="intron 8"
FT FT 6823
FT FT /tag=ac
FT FT /note="known polymorphic site"
FT FT 9106
FT FT /tag=ad
FT FT /note="known polymorphic site"
FT FT 9164.9209
FT FT /tag=ae
FT FT /note="exon 9"
FT FT 9207
FT FT /tag=af
FT FT /note="known polymorphic site"
FT FT 9210.10530
FT FT /tag=ag
FT FT /note="intron 9"
FT FT 9376
FT FT /tag=ah
FT FT /note="known polymorphic site"
FT FT 10531.10607
FT FT /tag=ai
FT FT /note="exon 10"
FT FT 10608.11597
FT FT /tag=aj
FT FT /note="intron 10"
FT FT 11384.11396
FT FT /tag=ak
FT FT /note="indefinite interval within intron 10"
FT FT 11598.15023
FT FT /tag=al
FT FT /note="exon 11"
FT FT 11908
FT FT /tag=am
FT FT /note="known polymorphic site"
FT FT 11994
FT FT /tag=an
FT FT /note="known polymorphic site"
FT FT 12952
FT FT /tag=ao
FT FT /note="known polymorphic site"
FT FT 13004
FT FT /tag=ap
FT FT /note="known polymorphic site"
FT FT 13009
FT FT /tag=aq
FT FT /note="known polymorphic site"
FT FT 13048
FT FT /tag=ar
FT FT /note="known polymorphic site"
FT FT 13238
FT FT /tag=as
FT FT /note="known polymorphic site"
FT FT 13448
FT FT /tag=at
FT FT /note="known polymorphic site"
FT FT 13539
FT FT /tag=au
FT FT /note="known polymorphic site"
FT FT 13951
FT FT /tag=av
FT FT /note="known polymorphic site"
FT FT 14041
FT FT /tag=aw
FT FT /note="known polymorphic site"
FT FT 14046
FT FT /tag=ax
FT FT /note="known polymorphic site"
FT FT 14475
FT FT /tag=ay
FT FT /note="known polymorphic site"
FT FT 14874
FT FT /tag=az
FT FT /note="known polymorphic site"
FT FT 14891
FT FT /tag=ba
FT FT /note="known polymorphic site"
FT FT 14966
```

```

FT      /*tag= bb
FT      /note= "known polymorphic site"
FT      15024. .15424
FT      /*tag= bc
FT      /note= "intron 11"
FT      15284
FT      /*tag= bd
FT      /note= "known polymorphic site"
FT      15425. .15511
FT      /*tag= be
FT      /note= "exon 12"
FT      15512. .15952
FT      /*tag= bf
FT      /note= "intron 12"
FT      15647. .15859
FT      /*tag= bg
FT      /note= "indefinite interval within intron 12"
FT      15953. .16126
FT      /*tag= bh
FT      /note= "exon 13"
FT      16077
FT      /*tag= bi
FT      /note= "known polymorphic site"
FT      16127. .16565
FT      /*tag= bj
FT      /note= "intron 13"
FT      16243
FT      /*tag= bk
FT      /note= "known polymorphic site"
FT      16370. .16382
FT      /*tag= bl

```

```

alignment_scores:
  Quality: 36.00      Length: 11
  Ratio: 3.600      Gaps: 0
  Percent Similarity: 90.909      Percent Identity: 63.636

```

```
alignment_block:

```

```
US-08-653-294-36 x T17517/rev ..

```

```
Align seg 1/1 to reverse of: T17517 from: 1 to: 24026

```

```

2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
11314 AGATTGCCATCAGAAACTGGTCTGATGTAC 11282

```

OM of: US-08-653-294-36 to: EST:\* out\_format : pfs

Date: Feb 8, 2000 6:23 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=Cpnl1.1/USPTO.spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.2  
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELET=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-36

Query length: 12

Database: EST:\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 7600.090000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_gss8:AQ051612	-	43.00	144.66	27.20	724	! A0851612 CpG1352B CpIOwAgDNA1 C
gb_est5:H72837	-	40.00	136.79	74.66	504	! H72837 y6c01.s1 Soares fetal
gb_est37:AI949481	+	40.00	136.04	82.19	547	! AI949481 wq10604.xl NCI_CGAP_KI
gb_est1:L33574	-	39.00	137.07	71.96	325	! L33574 BNAESTF350 Mustard flowe
gb_est21:AA983105	+	39.00	134.53	99.69	429	! AA983105 ua35b12.r1 Soares mos
gb_gss9:AQ137850	+	39.00	133.87	108.47	461	! AQ137850 HS_3058.B2_C09_MF CIT
gb_est44:AV396088	-	39.00	132.62	127.48	529	! AV396088 AV396088 Chlamydomonas
gb_est20:D25244	-	39.00	131.90	139.73	572	! D25244 HUMRES417B Clontech cata
gb_gss8:AQ057547	-	38.00	132.41	130.91	360	! AQ057547 CIT-HSP-235AG2.TR CIT-
gb_est37:AI943775	+	38.00	131.26	151.62	408	! AI943775 618045E09.xl 618 - inh
gb_gss3:BA74688	-	38.00	130.57	165.67	440	! BA74688 CIT-HSP-2034E11.TF CIT-H
gb_gss9:AQ122064	+	38.00	130.27	172.32	455	! AQ122064 HS_3077.B2_C10_MF CIT
gb_est19:AA802286	-	38.00	128.35	220.34	561	! AA802286 GM04141.5prime GM Dros
gb_est27:AI405570	-	38.00	128.17	225.42	572	! AI405570 GH25611.5prime GH Dros
gb_est27:AI402928	-	38.00	128.16	225.89	573	! AI402928 GH22268.5prime GH Dros
gb_est28:AI513785	-	38.00	128.16	225.89	573	! AI513785 GH26874.5prime GH Dros
gb_est25:AI297201	-	38.00	128.12	226.81	575	! AI297201 LP11454.5prime LP Dros
gb_est36:AI388173	-	38.00	128.12	226.81	575	! AI388173 GH18981.5prime GH Dros
gb_est38:AI517485	-	38.00	128.12	226.81	575	! AI517485 GH28449.5prime GH Dros
gb_est35:AI258224	-	38.00	127.69	239.83	603	! AI258224 LP01361.5prime LP Dros
gb_est25:AI258370	-	38.00	127.69	239.83	603	! AI258370 LP01550.5prime LP Dros
gb_est25:AI293190	-	38.00	127.67	240.30	604	! AI293190 GH16307.5prime LP Dros
gb_gss3:AQ815582	+	38.00	127.40	248.72	622	! AQ815582 HS_5348.A2_E06_SP6E RF
gb_est23:AI107515	-	38.00	127.17	256.25	638	! AI107515 GH05237.5prime GH Dros
gb_est23:AI135307	-	38.00	126.90	265.23	657	! AI135307 GH12986.5prime GH Dros
gb_gss1:AG000734	-	38.00	126.78	269.50	666	! AG000734 Homo sapiens genomic i
gb_gss6:AQ840156	-	38.00	126.58	276.64	681	! AQ840156 nbxb0053C09f CUGI Rice
gb_gss11:AQ327491	-	38.00	125.14	332.73	797	! AQ327491 nbxb0041M03r CUGI Rice
gb_est34:AV132998	-	37.00	124.97	94.22	181	! AV132998 AV132998 Mus musculus
gb_est1:N76762	-	37.00	128.81	207.75	355	! N76762 YH45C01.r1 Soares fetal
gb_est1:230048	-	37.00	128.81	207.75	355	! 230048 YH45C01r Atrium cdna libr
gb_gss7:AQ098311	+	37.00	128.33	220.86	374	! AQ098311 GSSTC05069 Trypanosom
gb_gss8:AQ075626	-	37.00	127.83	235.48	395	! AQ075626 CIT-HSP-235AC2.TR CIT-
gb_gss8:AQ044478	+	37.00	127.01	261.58	432	! AQ044478 CIT-HSP-238413.TF CIT
gb_est23:AI110273	+	37.00	126.97	263.00	434	! AI110273 BSMF524D19SK Brugla
gb_gss6:AQ869567	+	37.00	126.89	265.85	438	! AQ869567 nbmb0035D05f CUGI Rice
gb_gss8:AQ044345	+	37.00	126.80	268.70	442	! AQ044345 CIT-HSP-2329021.TF CIT
gb_gss6:AQ084684	+	37.00	126.74	270.84	445	! AQ084684 LMAJVL1.Lm23302.xl Lei
gb_gss10:AQ193772	+	37.00	126.70	272.27	447	! AQ193772 CIT-HSP-2383B10.TF CIT
gb_gss8:AQ043928	+	37.00	126.62	275.13	451	! AQ043928 CIT-HSP-2329013.TF CIT
gb_gss9:AQ111833	+	37.00	126.62	275.13	451	! AQ111833 CIT-HSP-2378D13.TF CIT
gb_gss8:AQ040090	+	37.00	126.56	277.28	454	! AQ040090 CIT-HSP-2333J23.TF CIT

gb\_gss8:AQ057253 + 37.00 126.54 278.00 455 ! AQ057253 CIT-HSP-2343E21.TF  
gb\_gss9:AQ114390 + 37.00 126.52 278.72 456 ! AQ114390 CIT-HSP-2377N1.TF C  
gb\_gss6:AQ870816 - 37.00 126.40 283.03 462 ! AQ870816 nbmb0041C06f CUGI R

seq\_name: gb\_gss6:AQ851612

seq\_documentation\_block: 724 bp DNA GSS 18-OCT-1999  
LOCUS AQ851612  
DEFINITION CpG1352B CpIOwAgDNA1 Cryptosporidium parvum genomic similar to SKB1  
homologue (negative regulator of mitosis) (regulator of Shk1, a  
p1(Cdc42/Rac)-activated kinase (PAK)), genomic survey sequence.  
ACCESSION AQ851612  
VERSION 1 GI:6063307  
KEYWORDS GSS.  
SOURCE Cryptosporidium parvum.  
ORGANISM Cryptosporidium parvum.  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;  
Cryptosporidiidae; Cryptosporidium.  
1 (bases 1 to 724)  
Strong, W.B. and Nelson, R.G.  
Cryptosporidium parvum GSS Project  
Unpublished (1997)  
On Sep 10, 1998 this sequence version replaced gi:3553959.  
Contact: Nelson, R. G.  
Depts. of Medicine & Pharmaceutical Chemistry  
San Francisco General Hospital-University of California, San  
Francisco  
Box 0811, San Francisco, CA 94143-0811, USA  
Tel: 415 206 8846  
Fax: 415 206 3353  
Email: malariagita.ucsf.edu  
For Annotation Data see http://medsfgh.ucsf.edu/id/CpTags/home.html  
Seq primer: T3  
Class: shotgun.

FEATURES  
source 1. 724  
Location/Qualifiers  
/organism="Cryptosporidium parvum"  
/strain="IOWA"  
/db\_xref="taxon:5807"  
/clone\_lib="CpIOwAgDNA1"  
/lab\_host="E. coli XL2 Blue MRF"  
/note="Vector: pBluescript II (SK-); Site1: EcoRV; C.  
parvum (IOWA isolate) genomic DNA was hydrodynamically  
sheared to produce fragments having a tight size  
distribution between 2-4 kb by Dr. Yvonne Thorntson of  
the Stanford DNA Sequencing and Technology Center  
(http://sequence.  
www.stanford.edu/group/techdev/shear.htm). The randomly  
sheared gDNA was chromatographed on Sephacryl S-400 to  
remove any small fragments and DNA eluting in the void  
volume was subcloned into an EcoR V-digested, alkaline  
phosphatase-treated pBluescript II (SK-) vector and  
transformed into E. coli strain XL2 Blue MRF.  
Recombinant clones from the first plating of the library  
were selected for sequence analysis using T3 and T7  
primers."

BASE COUNT 264 a 93 c 121 g 244 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 43.00 Length: 11  
Ratio: 4.300 Gaps: 0  
Percent Similarity: 90.909 Percent Identity: 90.909

alignment\_block:  
US-08-653-294-36 x AQ851612/rev ..

Align seg 1/1 to reverse of: AQ851612 from: 1 to: 724

1 TyrArgLeuAlaIleArgArgIleAlaLeuArg 11  
|||||||  
552 TATAGGCTCTAATAAGACGTACGTTGAGG 520

```

seq_name: gb_est5:H72837
seq_documentation_block: 504 bp mRNA EST 27-OCT-1995
LOCUS H72837
DEFINITION Ys06c01.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:213984 3', mRNA sequence.
ACCESSION H72837
VERSION H72837.1 GI:1044653
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 504)
AUTHORS Hillier,L., Lennan,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevisan,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.,
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 9704478
COMMENT On May 8, 1995 this sequence version replaced gi:799648.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 528
High quality sequence stops: 380
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 528 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 380.
FEATURES
Location/Qualifiers
1..504
/organism="Homo sapiens"
/db_xref="GDB:3779288"
/db_xref="taxon:9606"
/clone="IMAGE:213984"
/clone_lib="Soares fetal liver spleen 1NFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGGAAGAAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 120 a 103 c 125 g 147 t
ORIGIN
alignment_scores:
Quality: 40.00 Length: 12
Ratio: 4.000 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 58.333
alignment_block:
US-08-653-294-36 x H72837/rev
Align seg 1/1 to reverse of: H72837 from: 1 to: 504

```

```

1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
289 TACCCTAGCCCTGAARAGGTAAACCTCGATAT 254

seq_name: gb_est37:AI949481
seq_documentation_block: 547 bp mRNA EST 06-SEP-1999
LOCUS AI949481
DEFINITION wq10g04.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2470902 3',
mRNA sequence.
ACCESSION AI949481
VERSION AI949481.1 GI:5741713
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247088.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

```

```

Seq primer: -400P from Gibco
High quality sequence stop: 454.
FEATURES
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2470902"
/clone_lib="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 134 a 92 c 109 g 212 t
ORIGIN
alignment_scores:
Quality: 40.00 Length: 12
Ratio: 4.000 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667
alignment_block:
US-08-653-294-36 x AI949481
Align seg 1/1 to: AI949481 from: 1 to: 547
1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12
|||||
484 TACAGGCTTCCTATTAGAGATTAGCATTCATGTGG 519

```

WashU-HHMI Mouse EST Project  
Contact: Maria M/Mouse EST Project

RECEIVED  
2000000  
COMMENT: Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3058 row: F column: 18  
Class: BAC ends  
High quality sequence stop: 461.

## FEATURES

source  
1. .461  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate=3058 Col=18 Row=F"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 132 a 125 c 102 g 102 t  
ORIGIN

alignment\_scores:  
Quality: 39.00 Length: 12  
Ratio: 3.900 Gaps: 0  
Percent Similarity: 83.333 Percent Identity: 66.667

## alignment\_block:

US-08-653-294-36 x AQ137850 ..

Align seg 1/1 to: AQ137850 from: 1 to: 461

1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12  
18 CATCGACTCACTATAAGCGGAATTCGCCAGGTAC 53

seq\_name: gb\_est44:AV396088

seq\_documentation\_block: 529 bp mRNA EST 09-DEC-1999  
LOCUS AV396088  
DEFINITION AV396088 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii  
cDNA clone CL56H04\_r, mRNA sequence.

ACCESSION AV396088

VERSION AV396088.1 GI:6550304

## KEYWORDS

SOURCE

## ORGANISM

Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.

## REFERENCE

1 (bases 1 to 529)

Asamizu E., Nakamura Y., Sato S., Fukuzawa H. and Tabata S.  
A Large Scale Structural Analysis of cDNAs in a Unicellular Green  
Alga, Chlamydomonas reinhardtii. I. Generation of 3451  
non-redundant Expressed Sequence Tags

## JOURNAL

DNA Res. (1999) In press

On May 18, 1998 this sequence version replaced gi:3137395.

## COMMENT

Contact: Yasukazu Nakamura

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: ynakamuka@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

source  
1. .529  
Location/Qualifiers  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="CL56H04\_r"  
/clone\_lib="Chlamydomonas reinhardtii C9"  
/dev\_stage="photoautotrophic growth"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

## BASE COUNT

106 a 179 c 140 g 104 t

## ORIGIN

alignment\_scores:  
Quality: 39.00 Length: 10  
Ratio: 3.900 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 70.000

## alignment\_block:

US-08-653-294-36 x AV396088/rev ..

Align seg 1/1 to reverse of: AV396088 from: 1 to: 529

2 ArgLeuAlaIleArgArgIleAlaLeuArg 11

|||||:|||||:|||||:|||||:|||||

520 CGACTTCGGTCCGAGGTCGCGGTGCGA 491

seq\_name: gb\_est20:D25244

seq\_documentation\_block: 572 bp mRNA EST 12-MAR-1998  
LOCUS D25244  
DEFINITION HUMRES417B Clontech catalog #CLHL1076b Homo sapiens cDNA clone  
RES4-17, mRNA sequence.

ACCESSION D25244

VERSION D25244.1 GI:434739

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 572)

Ikeda J., Hadano S., Nagayama T., Tomiyasu H., Wakasa K. and  
Ikeda J.

Isolation and characterization of 21 novel expressed DNA sequences  
from the distal region of human chromosome 4p

Genomics 22 (2), 302-312 (1994)

95104839

COMMENT

On May 9, 1995 this sequence version replaced gi:804176.

Contact: Yoshikazu Ishida

Ikeda Genosphere project/ERATO/JRDC

Tokai University School of Medicine

Bohsaidai, Isehara, Kanagawa 259-1193, Japan

Email: shinjien@med.u-tokai.ac.jp.

## FEATURES

Location/Qualifiers

1. .572

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="4p15-4p16.1"

/clone="RES4-17"

/clone\_lib="Clontech catalog #CLHL1076b"

/tissue\_type="caudate"

/dev\_stage="20-year-old adult"

BASE COUNT 190 a 87 c 121 g 156 t 18 others

## ORIGIN

alignment\_scores:  
Quality: 39.00 Length: 12  
Ratio: 3.900 Gaps: 0  
Percent Similarity: 83.333 Percent Identity: 58.333

## alignment\_block:

US-08-653-294-36 x D25244/rev ..

Align seg 1/1 to reverse of: D25244 from: 1 to: 572

1 TyrArgLeuAlaIleArgArgIleAlaLeuArgTyr 12

|||||:|||||:|||||:|||||:|||||

309 TTTTCCCTGTCGCAATAGAGTAAGTTTGATAC 274

seq\_name: gb\_gss8:AQ075747

## seq\_documentation\_block:



LOCUS AQ075747 360 bp DNA GSS 20-AUG-1998  
DEFINITION CIT-HSP-2354G2.TR CIT-HSP Homo sapiens genomic clone 2354G2,  
genomic survey sequence.  
ACCESSION AQ075747  
VERSION AQ075747.1 GI:3437403  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 360)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shizuwa,H., Simon,M. and  
Venter,J.C.  
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building  
JOURNAL Map Building (1998)  
COMMENT Other\_GSSs: CIT-HSP-2354G2.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.  
FEATURES  
source  
1..360  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="2354G2"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOAC11; Site\_1: HindIII; Site\_2:  
HindIII"  
BASE COUNT 76 a 120 c 78 g 85 t 1 others  
ORIGIN  
  
alignment\_scores:  
Quality: 38.00 Length: 11  
Ratio: 3.800 Gaps: 0  
Percent Similarity: 90.909 Percent Identity: 72.727  
  
alignment\_block:  
US-08-653-294-36 x AQ075747/rev ..  
Align seg 1/1 to reverse of: AQ075747 from: 1 to: 360  
  
2 ArgLeuAlaIleArgArgIleAlaLeuArgTyr 12  
|||||:|||||:|||||:|||||  
55 CGACTCACTATAAGCGCAATTAGNGCTCGGTAC 23  
  
seq\_name: gb\_est37:AI943775  
  
seq\_documentation\_block:  
LOCUS AI943775 408 bp mRNA EST 09-AUG-1999  
DEFINITION 618045E09.xl 618 - Inbred rassel cdna Library zea mays CDNA, mRNA  
sequence.  
ACCESSION AI943775  
VERSION AI943775.1 GI:5713783  
KEYWORDS EST.  
SOURCE zea mays.  
ORGANISM zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Zea.  
REFERENCE 1 (bases 1 to 408)

AUTHORS Walbot,V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
JOURNAL Unpublished (1999)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135666.  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 618045 row: E column: 09.  
FEATURES  
Location/Qualifiers  
1..408  
/organism="Zea mays"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/clone\_lib="618 - Inbred Tassel cDNA Library"  
/tissue\_type="tassel"  
/dev\_stage="tassel"  
/lab\_host="XLOLR"  
/note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);  
Inbred tassel library from Schmidt lab"  
BASE COUNT 80 a 120 c 136 g 72 t  
ORIGIN  
  
alignment\_scores:  
Quality: 38.00 Length: 10  
Ratio: 3.800 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 70.000  
  
alignment\_block:  
US-08-653-294-36 x AI943775 ..  
Align seg 1/1 to: AI943775 from: 1 to: 408  
  
2 ArgLeuAlaIleArgArgIleAlaLeuArg 11  
|||||:|||||:|||||:|||||  
283 CGGTAGCCCTCGGAGGTAGCCCTCGCG 312  
  
seq\_name: gb\_gss3:B74688  
  
seq\_documentation\_block:  
LOCUS B74688 440 bp DNA GSS 24-OCT-1998  
DEFINITION CIT-HSP-2034E11.TF CIT-HSP Homo sapiens genomic clone 2034E11,  
genomic survey sequence.  
ACCESSION B74688  
VERSION B74688.1 GI:2770375  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 440)  
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuwa,H.,  
Simon,M. and Venter,J.C.  
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: CIT-HSP-2034E11.TP  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html

Seq primer: M13-21

Class: BAC ends.

FEATURES  
source

Location/Qualifiers

1. .440

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2034E11"

/clone\_lib="CIT-HSP"

/sex="Male"

/cell\_type="Sperm"

/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:

HindIII"

136 a 88 c 78 g 138 t

BASE COUNT  
ORIGIN

alignment\_scores:

Quality: 38.00 Length: 12

Ratio: 4.222 Gaps: 0

Percent Similarity: 75.000 Percent Identity: 66.667

alignment\_block:

US-08-653-294-36 x B74688 ..

Align seg 1/1 to: B74688 from: 1 to: 440

1 TyrArgLeuAlaileArgArgIleAlaLeuArgTyr 12

|||||:|||||

2 TACGCACCTACTATACGGCGAATCGCTCGGTAC 37

|||||

seq\_name: gb\_gss9:AQ122064

seq\_documentation\_block:

LOCUS AQ122064 455 bp DNA GSS 22-SEP-1998

DEFINITION HS\_3077\_B2\_C10\_MF CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=3077 Col=20 Row=F, genomic survey

sequence.

ACCESSION AQ122064

VERSION AQ122064.1 GI:3499230

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 455)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380389

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3077 row: F column: 20

Class: BAC ends

High quality sequence stop: 455.

Location/Qualifiers

1. .455

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=3077 Col=20 Row=F"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in

E-Coli DH10B"

BASE COUNT 159 a 81 c 92 g 122 t 1 others  
ORIGIN

alignment\_scores:

Quality: 38.00 Length: 10

Ratio: 3.800 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 80.000

alignment\_block:

US-08-653-294-36 x AQ122064 ..

Align seg 1/1 to: AQ122064 from: 1 to: 455

2 ArgLeuAlaileArgArgIleAlaLeuArg 11

|||||:|||||

17 CGACTGCTATACGTCGAATCACTATACGG 46

|||||

seq\_name: gb\_est19:AA802286

seq\_documentation\_block:

LOCUS AA802286 561 bp mRNA EST 25-NOV-1998

DEFINITION GM04141.5prime GM Drosophila melanogaster ovary Bluescript

Drosophila melanogaster cDNA clone GM04141 5prime, mRNA sequence.

ACCESSION AA802286

VERSION AA802286.1 GI:2871405

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 561)

Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,

Brokstein,P., Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1402170.

Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology

University of California Berkeley

539 LSA, Berkeley, CA 94720-3200, USA

Fax: 510 643 9947

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

Plate: 41 row: D column: 5

High quality sequence stop: 411.

Location/Qualifiers

1. .561

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GM04141"

/clone\_lib="GM Drosophila melanogaster ovary Bluescript"

/sex="female"

/dev\_stage="newly enclosed females: germarium-stage 6"

/lab\_host="SOLR"

/notes="Organ: ovary; Vector: Bluescript SK; Site\_1: EcoRI;

Site\_2: XhoI; Constructed using Stratagene ZAP-cDNA

Synthesis kit. Oligo dT-primed and directionally cloned at

EcoRI and XhoI in Bluescript SK(+/-)"

153 a 166 c 141 g 101 t

BASE COUNT  
ORIGIN

alignment\_scores:

Quality: 38.00 Length: 12

Ratio: 3.455- Gaps: 0

Percent Similarity: 91.667 Percent Identity: 58.333

alignment\_block:

US-08-653-294-36 x AA802286/rev ..

Align seg 1/1 to reverse of: AA802286 from: 1 to: 561

SOURCE fruit fly.

**THIS PAGE BLANK (UPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 04:05:45 ; Search time 133.56 Seconds  
(without alignments)  
1.773 Million cell updates/sec

Title: US-08-653-294-37  
Perfect score: 51  
Sequence: 1 RVDLRLRGY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	88.2	274	1 P80911	Consensus sequence
2	44	86.3	10	1 W07517	T-cell modulating
3	44	86.3	17	1 R71439	Human HLA-A2-(62-8
4	44	86.3	17	1 R71441	Human MHC 1 and HL
5	44	86.3	24	1 R84876	HLA A2.1 61-84 1
6	44	86.3	24	1 R84875	HLA A2.1 61-84 imm
7	44	86.3	25	1 R41222	Peptide fragment o
8	44	86.3	25	1 R83074	HLA-A2 CTL modul
9	44	86.3	25	1 R95420	HLA-A2.1 60-84. Co
10	44	86.3	412	1 W68385	Chimeric HLA-A2.1/
11	40	78.4	22	1 R44112	HLA epitope. HIV G
12	40	78.4	25	1 R41206	Peptide fragment o
13	40	78.4	25	1 R83072	HLA-Bw46 CTL modul
14	40	78.4	25	1 R95418	HLA-Bw46.60-84. Co
15	40	78.4	45	1 R71629	HLA-alpha-1. Use o
16	40	78.4	366	1 R12455	HLA-C exon Cb-1. H
17	40	78.4	366	1 R12456	HLA-C exon Cb-2. H
18	40	78.4	366	1 Y07033	Breast cancer asso
19	39	76.5	10	1 R41211	Peptide fragment o
20	39	76.5	10	1 W07521	T-cell modulating
21	39	76.5	25	1 P90513	Alpha-1 domain of
22	39	76.5	25	1 R08111	Example of cell su
23	39	76.5	25	1 R69618	MHC class I-derive
24	38	74.5	10	1 R41210	Peptide fragment o
25	38	74.5	10	1 W07516	T-cell modulating
26	38	74.5	17	1 R71431	Human MHC 1 alpha
27	38	74.5	5069	1 W52846	A. mediterranei ri
28	37	72.5	9	1 R84869	H2Kb 75-83 immunog
29	37	72.5	17	1 R71430	Human MHC 1 alpha
30	36	70.6	10	1 W07524	T-cell modulating
31	36	70.6	13	1 W29421	Glucose transport
32	36	70.6	17	1 R71440	Human MHC 1 and HL
33	36	70.6	17	1 R71425	Human MHC 1 alpha
34	36	70.6	17	1 R71426	Human MHC 1 alpha

Human MHC 1 alpha  
MHC peptide repeat  
MHC peptide repeat  
Glucose transport  
Peptide membrane b  
Human MHC 1 alpha  
Human MHC 1 alpha  
MHC class I-derive  
MHC-I peptide Dk-(  
Human MHC 1 alpha  
MHC peptide having

## ALIGNMENTS

## RESULT 1

P80911  
ID P80911 standard; protein; 274 AA.  
AC P80911;  
DT 18-SEP-1990 (first entry)  
DE Consensus sequence of peptides which constitute the alpha-1, alpha-2 and  
DE alpha-3 regions of a class I HLA molecule  
KW HLA-A2 epitopes; extracellular domains alpha-1, alpha-2 and alpha-3.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..90  
FT region /note="alpha-1 region"  
FT region 91..180  
FT region /note="alpha-2 region"  
FT region 181..272  
FT region /note="alpha-3 region"  
PN W0805784-A.  
PD 11-AUG-1988.  
PF 27-FEB-1988; U00245.  
PR 30-JAN-1987; US-008846.  
PA (STRD) Leland Stanford Jr Univ.  
PI Krensky AM, Parham P, Clayberger C;  
DR WPI; 88-235147/33.  
PT New peptide corresp. to major histocompatibility antigen regions -  
PT used for modulating cytotoxic T-lymphocyte activity in e.g.  
PT transplants or auto-immune diseases  
PS Example 9; Fig 4; 60pp; English.  
CC The consensus sequence is derived from a total of 23 HLA-A,B,C sequences.  
CC The protein sequences in the three extracellular domains (alpha-1,  
CC alpha-2 and alpha-3) are shown. The example concerned the effect of  
CC peptides from different HLA-A2 epitopes on cytotoxicity of target cells by  
CC CTL of different specificities.  
SQ Sequence 274 AA;

Query Match 88.2%; Score 45; DB 1; Length 274;  
Best Local Similarity 90.0%; Pred. No. 0.1;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10  
1 | | | | | | | | | |  
DB 75 RVDLRLRGY 84

## RESULT 2

W07517  
ID W07517 standard; peptide; 10 AA.  
AC W07517;  
DT 04-AUG-1997 (first entry)  
DE T-cell modulating peptide #6.  
KW T-cell modulator; autoimmune disease; tissue destruction; alapha-domain;  
KW mammal; major histocompatibility complex; MHC class I; antigen; perforin;  
KW insulin-dependent diabetes mellitus; multiple sclerosis; inflammation;  
KW rheumatoid arthritis; psoriasis; pemphigus vulgaris; Sjogren's disease;  
KW thyroid disease; Hashimoto's thyroiditis; myasthenia gravis; granzyme;  
KW autologous target cell; cytokine release; T cell activation; therapy.  
OS Synthetic.

PN W09635443-A1.  
 PD 14-NOV-1996.  
 PF 05-APR-1996; U04710.  
 PR 12-MAY-1995; US-440504.  
 PA (SANG-) SANGSTAT MEDICAL CORP.  
 PI Buelow R;  
 DR WPI: 96-518410/51.  
 PT Treatment of auto-immune disease by admin. of peptide(s) corresp. to  
 PT major histocompatibility complex antigens - esp. for delaying onset  
 PT of clinical symptoms of insulin dependent diabetes by modulating T  
 PT cell mediated attack on target cells  
 PS Claim 7; Page 20; 24pp; English.  
 CC W07512-W07518 represent T-cell modulating peptides that can be used in  
 CC the method of the invention. These sequences are based on a portion of  
 CC the generic peptide corresponding to residues 70-91 of the alpha 1-domain  
 CC of the major histocompatibility complex (MHC) class I antigen (see  
 CC W07510). The method is for affecting the course of an autoimmune disease  
 CC involving T-cell mediated destruction of tissue in mammals. These  
 CC peptides are used especially to treat insulin-dependent diabetes  
 CC mellitus, preferably being administered during the pre-clinical stage to  
 CC delay onset of the disease. Other diseases that can be treated are  
 CC multiple sclerosis, rheumatoid arthritis, psoriasis, pemphigus vulgaris,  
 CC Sjogren's disease, thyroid disease, Hashimoto's thyroiditis, myasthenia  
 CC gravis, etc. The peptides modulate T-cell mediated attack on autologous  
 CC target cells, and may also reduce inflammation, swelling, and release of  
 CC cytokines, perforins, granzymes etc. associated with T cell activation.  
 SQ Sequence 10 AA;

Query Match 86.3%; Score 44; DB 1; Length 10;  
 Best Local Similarity 90.0%; Pred. No. 0.0047;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10  
 ||||| ||  
 DB 1 RVDLRLRGY 10

RESULT 3  
 R71439 ID R71439 standard; peptide; 17 AA.  
 AC R71439;  
 DT 12-OCT-1995 (first entry)  
 DE Human HLA-A2-(62-85) antigen derived peptide.  
 KW Human HLA-A2-(62-85) antigen derived peptide; cell receptor;  
 KW interaction modulation; arthritis; neoplasias; lupus erythematosus.  
 OS Homo sapiens.  
 PN W09505189-A.  
 PD 23-FEB-1995.  
 PF 12-AUG-1993; U09189.  
 PR 12-AUG-1993; US-105416.  
 PA (REGC) UNIV CALIFORNIA.  
 PI Goldstein A, Goodenow RS, Olsson L;  
 DR WPI: 95-098577/13.  
 PT Regulating cell surface receptor response - by modulating  
 PT interaction between MHC class I antigen and the cell surface  
 PT receptor  
 PS Example 4; Page 45; 103pp; English.  
 CC R71439-R71443 are human major histocompatibility complex class I  
 CC (MHC 1) alpha 1 domain and HLA derived peptides and fusion peptides.  
 CC They were used to modulate interactions between MHC 1/HLA and cell  
 CC surface receptors. Via competitive inhibition the peptides diminish  
 CC the receptors response, this feature may be useful for the treatment  
 CC of neoplasias, lupus erythematosus and arthritis.  
 SQ Sequence 17 AA;

Query Match 86.3%; Score 44; DB 1; Length 17;  
 Best Local Similarity 90.0%; Pred. No. 0.0083;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10  
 ||||| |||||  
 DB 7 RVDLRLRGY 16

RESULT 5  
 R84876 ID R84876 standard; peptide; 24 AA.  
 AC R84876;  
 DT 30-MAY-1996 (first entry)  
 DE HLA AW68.1 61-84 immunogenic peptide.  
 KW Antigen; epitope; cell mediated; immune specific; cancer;  
 KW infection; infestation; mucin-1; MUC-1; tumour; HLA AW68.1 61-84;  
 KW immunogenic peptide.  
 OS Synthetic.  
 PN W09527505-A1.  
 PD 19-OCT-1995.  
 PF 12-APR-1995; U04540.  
 PR 12-APR-1994; US-229606.  
 PA (BIOM-) BIOMIRA INC.  
 PI Ding L, Koganty RR, Longenecker BM, Reddish MA;  
 DR WPI: 95-373528/48.  
 PT New cell-mediated immune-specific immunogenic compns. - used in  
 PT prophylaxis and treatment of cancer, microbial infections, viral  
 PT infections and parasitic infestations  
 PS Example 1; Page 62; 141pp; English.  
 CC A claimed cell-mediated immune (CMI)-specific immunogenic compsn.  
 CC comprises a conjugate of a primary antigen bearing a primary epitope,  
 CC with an immunomodulatory peptide (IP), i.e. the present peptide.  
 CC The IP comprises an alloepitope moiety of at least 5 amino acids,  
 CC whose sequence corresponds to a polymorphic region of a MHC  
 CC encoded polymorphic Class I or II antigen. The compsn. can be  
 CC used to elicit a CMI-specific response which is prophylactic, or  
 CC therapeutic for, e.g. microbial and viral infections, parasitic  
 CC infestations and cancer, partic. MUC-1 expressing tumour cells  
 CC when the present peptide is the IP, and a MUC-1 epitope is the

Query Match 86.3%; Score 44; DB 1; Length 17;  
 Best Local Similarity 90.0%; Pred. No. 0.0083;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10  
 ||||| |||||

DB 7 RVDLRLRGY 16

RESULT 4  
 R71441 ID R71441 standard; peptide; 17 AA.  
 AC R71441;  
 DT 12-OCT-1995 (first entry)  
 DE Human MHC 1 and HLA fusion peptide DK-(69-76)HLA-A2-(77-85).  
 KW Major histocompatibility complex class I; MHC 1; cell receptor;  
 KW alpha 1 domain; fusion peptide DK-(69-76)HLA-A2-(77-85);  
 KW interaction modulation; arthritis; neoplasias; lupus erythematosus.  
 OS Homo sapiens.  
 PN W09505189-A.  
 PD 23-FEB-1995.  
 PF 12-AUG-1994; U09189.  
 PR 12-AUG-1993; US-105416.  
 PA (REGC) UNIV CALIFORNIA.  
 PI Goldstein A, Goodenow RS, Olsson L;  
 DR WPI: 95-098577/13.  
 PT Regulating cell surface receptor response - by modulating  
 PT interaction between MHC class I antigen and the cell surface  
 PT receptor  
 PS Example 4; Page 45; 103pp; English.  
 CC R71439-R71443 are human major histocompatibility complex class I  
 CC (MHC 1) alpha 1 domain and HLA derived peptides and fusion peptides.  
 CC They were used to modulate interactions between MHC 1/HLA and cell  
 CC surface receptors. Via competitive inhibition the peptides diminish  
 CC the receptors response, this feature may be useful for the treatment  
 CC of neoplasias, lupus erythematosus and arthritis.  
 SQ Sequence 17 AA;

Query Match 86.3%; Score 44; DB 1; Length 17;  
 Best Local Similarity 90.0%; Pred. No. 0.0083;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10  
 ||||| |||||  
 DB 7 RVDLRLRGY 16

RESULT 5  
 R84876 ID R84876 standard; peptide; 24 AA.  
 AC R84876;  
 DT 30-MAY-1996 (first entry)  
 DE HLA AW68.1 61-84 immunogenic peptide.  
 KW Antigen; epitope; cell mediated; immune specific; cancer;  
 KW infection; infestation; mucin-1; MUC-1; tumour; HLA AW68.1 61-84;  
 KW immunogenic peptide.  
 OS Synthetic.  
 PN W09527505-A1.  
 PD 19-OCT-1995.  
 PF 12-APR-1995; U04540.  
 PR 12-APR-1994; US-229606.  
 PA (BIOM-) BIOMIRA INC.  
 PI Ding L, Koganty RR, Longenecker BM, Reddish MA;  
 DR WPI: 95-373528/48.  
 PT New cell-mediated immune-specific immunogenic compns. - used in  
 PT prophylaxis and treatment of cancer, microbial infections, viral  
 PT infections and parasitic infestations  
 PS Example 1; Page 62; 141pp; English.  
 CC A claimed cell-mediated immune (CMI)-specific immunogenic compsn.  
 CC comprises a conjugate of a primary antigen bearing a primary epitope,  
 CC with an immunomodulatory peptide (IP), i.e. the present peptide.  
 CC The IP comprises an alloepitope moiety of at least 5 amino acids,  
 CC whose sequence corresponds to a polymorphic region of a MHC  
 CC encoded polymorphic Class I or II antigen. The compsn. can be  
 CC used to elicit a CMI-specific response which is prophylactic, or  
 CC therapeutic for, e.g. microbial and viral infections, parasitic  
 CC infestations and cancer, partic. MUC-1 expressing tumour cells  
 CC when the present peptide is the IP, and a MUC-1 epitope is the

CC primary epitope.  
SQ Sequence 24 AA;

Query Match 86.3%; Score 44; DB 1; Length 24;  
Best Local Similarity 90.0%; Pred. No. 0.012;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTGRY 10  
| | | | | | | | | |  
DB 15 RVDLGLTGRY 24

## RESULT 6

ID R84875 standard; peptide; 24 AA.  
AC R84875;  
DT 30-MAY-1996 (first entry)  
DE HLA A2.1 61-84 immunogenic peptide.  
KW Antigen: epitope; cell mediated; immune specific; cancer;  
KW Infection: infestation; mucin-1; MUC-1; tumour; HLA A2.1 61-84;  
KW immunogenic peptide.  
OS Synthetic.  
PN WO9527505-A1.  
PD 19-OCT-1995.  
PF 12-APR-1995; U04540.  
PR 12-APR-1994; US-229606.  
PA (BIOM-) BIOMIRA INC.  
PI Ding L, Koganty RR, Longenecker BM, Reddish MA;  
DR WPI: 95-373528/48.  
PT New cell-mediated immune-specific immunogenic compsns. - used in  
PT prophylaxis and treatment of cancer, microbial infections, viral  
PT infections and parasitic infestations  
PS Example 1; Page 62; 14pp; English.  
CC A claimed cell-mediated immune (CMI)-specific immunogenic compsn.  
CC comprises a conjugate of a primary antigen bearing a primary epitope,  
CC with an immunomodulatory peptide (IP), i.e. the present peptide.  
CC The IP comprises an alloptide moiety of at least 5 amino acids,  
CC whose sequence corresponds to a polymorphic region of a MHC  
CC encoded polymorphic Class I or II antigen. The compsn. can be  
CC used to elicit a CMI-specific response which is prophylactic, or  
CC therapeutic for, e.g. microbial and viral infections, parasitic  
CC infestations and cancer, partic. MUC-1 expressing tumour cells  
CC when the present peptide is the IP, and a MUC-1 epitope is the  
CC primary epitope.  
SQ Sequence 24 AA;

Query Match 86.3%; Score 44; DB 1; Length 24;  
Best Local Similarity 90.0%; Pred. No. 0.012;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTGRY 10  
| | | | | | | | | |  
DB 15 RVDLGLTGRY 24

## RESULT 7

ID R41222 standard; peptide; 25 AA.  
AC R41222;  
DT 15-MAR-1994 (first entry)  
DE Peptide fragment of HLA-A2.1 antigen.  
KW Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;  
KW parasitic disease; cytotoxic T lymphocyte; modulation.  
OS Synthetic.  
PN WO9317699-A.  
PD 16-SEP-1993.  
PF 25-FEB-1993; U01758.  
PR 02-MAR-1992; US-844716.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger CA, Krensky AM;  
DR WPI: 93-303134/38.

PT New peptide(s) based on Class I HLA-antigen domains - used for  
PT modulating cytotoxic T-lymphocyte activity towards targets  
PS Example 13; Page 39; 61pp; English.  
CC The peptide is used to modulate cytotoxic T-lymphocyte (CTL)  
CC activity, either by inhibition or stimulation. It can be used for  
CC inhibiting CTL toxicity in transplantations, for inducing CTL  
CC activity in parasitic diseases and neoplasia and in studies on viral  
CC infection. The peptide can also be used for identifying CTLs which  
CC bind to it and removing subsets of CTLs from a T-cell composition.  
CC This peptide is derived from the HLA-A2.1 antigen and corresponds to  
CC the amino acid positions 60-84 of that antigen.  
SQ Sequence 25 AA;

Query Match 86.3%; Score 44; DB 1; Length 25;  
Best Local Similarity 90.0%; Pred. No. 0.013;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTGRY 10  
| | | | | | | | | |  
DB 16 RVDLGLTGRY 25

## RESULT 8

ID R83074 standard; peptide; 25 AA.  
AC R83074;  
DT 16-MAY-1996 (first entry)  
DE HLA-A2 CTL modulating peptide (A2.1.60-84).  
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
KW class I MHC; HLA-A2.  
OS Synthetic.  
PN WO9528979-A1.  
PD 12-OCT-1995.  
PF 05-APR-1995; U04349.  
PR 05-APR-1994; US-222851.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM, Parham P;  
DR WPI: 95-358582/46.  
PT Extension of acceptance period of transplants from MHC unmatched  
PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
PT host  
PS Example 13; Page 32; 80pp; English.  
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
CC class I major histocompatibility complex (MHC) antigens. This sequence  
CC corresponds to residues 60-84 of the alpha-1 domain of the class  
CC I MHC HLA-A2. These sequences can be used to extend the period of  
CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
CC The peptides are administered to a patient in conjunction with a  
CC subtherapeutic amount of an immunosuppressant. This is administered to  
CC the patient for a limited period of time (compared to the lifetime  
CC administration for current treatments). The peptides particularly  
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
CC of the patient.  
SQ Sequence 25 AA;

Query Match 86.3%; Score 44; DB 1; Length 25;  
Best Local Similarity 90.0%; Pred. No. 0.013;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTGRY 10  
| | | | | | | | | |  
DB 16 RVDLGLTGRY 25

## RESULT 9

ID R95420 standard; peptide; 25 AA.  
AC R95420;  
DT 12-NOV-1996 (first entry)  
DE HLA-A2.1.60-84.

KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
 KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
 KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
 KW Cytolysis; antigen presenting cell.  
 OS Synthetic.  
 PN WO9513288-A1.  
 PD 18-MAY-1995.  
 PE 10-NOV-1994; U12985.  
 PR 10-NOV-1993; US-150493.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AW;  
 DR WPI; 95-194027/25.  
 PT Compens. comprising lymphoid surface membrane proteins - which may  
 PT inhibit cytolytic activity and differentiation of CTLs.  
 PS Example; Page 9; 29pp; English.  
 CC R95413, and R95415-R95431 represent palindromes and fragments of  
 CC human-leucocyte-associated antigens. This sequence represents the  
 CC HLA-A2.1.60-84. These sequences can be used to isolate the protein p74  
 CC from a T-cell lysate. p74 is a T-cell surface membrane protein  
 CC associated with T-cell activation in mammalian T-cells, and is also  
 CC immunologically cross reactive with the heat shock protein Hsc70. p74 is  
 CC found in a limited number of cell types, but is particularly expressed on  
 CC B and T cells. p74 can be isolated by lysis of a suitable cell with an  
 CC amphoteric detergent, and then passed through an affinity column  
 CC containing a covalently bound HLA-B2702 palindromic peptide.  
 CC Compositions comprising the extracellular fragment of p74 combined with  
 CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
 CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
 CC compounds can be screened for their effect on the cytolytic activity of  
 CC T-cells, by combining them with the extracellular portion of p74 and  
 CC determining the amount of binding between the candidate compound and p74.  
 CC Modulation of CTL activity can be inhibited in a cellular composition  
 CC containing T-cells and antigen presenting cells (APCs), by adding to the  
 CC mix the extracellular portion of p74, in an amount sufficient to compete  
 CC with p74 for the binding of the p74 ligand.  
 SQ Sequence 25 AA;

Query Match 86.3%; Score 44; DB 1; Length 25;  
 Best Local Similarity 90.0%; Pred. NO. 0.013;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTIRGY 10  
 ||||| |||||  
 DB 16 RVDLGLTIRGY 25

## RESULT 10

W68385  
 ID W68385 standard; Protein; 412 AA.  
 AC W68385;  
 DT 14-OCT-1998 (first entry)  
 DE Chimeric HLA-A2.1/beta-2 microglobulin protein.  
 KW Antigen; major histocompatibility complex; MHC; lymphocyte; detection;  
 KW Immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;  
 KW viral infection; chimeric; beta-2 microglobulin.  
 OS Synthetic.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 1..302  
 FT Region /note- "from HLA-A2.1 protein"  
 FT Region 303..312  
 FT Region /note- "flexible linker"  
 FT Domain 313..412  
 FT Domain /note- "from human beta-2 microglobulin"  
 PN WO9744667-A2.  
 PD 27-NOV-1997.  
 PE 21-MAY-1997; F00892.  
 PR 21-MAY-1996; US-651925.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (INSP) INST PASTEUR.  
 PI Abastado J, Kourilsky P, Langlade-Demoyen P, Lone Y;  
 DR WPI; 98-018653/02.

DR N-PSDB; V30457.  
 PT Detection, purification and elimination of antigen-specific  
 PT lymphocytes - for producing cytotoxic T cells for immuno-therapy of  
 PT cancers and viral infection  
 PS Example 2; Fig 10; 222pp; French.  
 CC Detection of antigen-specific lymphocytes comprises forming a complex  
 CC between antigenic peptides (see W68301-W68384 for examples) and  
 CC recombinantly produced major histocompatibility complex (MHC) molecules,  
 CC immobilising the complex and binding samples containing the  
 CC antigen-specific lymphocytes. Expression of the MHC molecule in a cell  
 CC is detected by antibodies against the MHC molecule or by an antibody to  
 CC a chimeric MHC-marker protein, e.g. an MHC/beta-2-microglobulin chimeric  
 CC protein. This sequence is an example of the chimeric protein and  
 CC comprises the first 3 domains of the HLA-A2.1 heavy chain linked to human  
 CC beta-2 microglobulin via a flexible linker. The method is also used to  
 CC detect and quantify tumour-specific T-cells and to generate CTC for  
 CC specific killing of tumour cells (solid tumours, leukaemia or lymphoma)  
 CC by injection into a human or animal, but also for treating viral  
 CC infections.  
 SQ Sequence 412 AA;

Query Match 86.3%; Score 44; DB 1; Length 412;  
 Best Local Similarity 90.0%; Pred. NO. 0.25;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTIRGY 10  
 ||||| |||||  
 DB 99 RVDLGLTIRGY 108

## RESULT 11

R44112  
 ID R44112 standard; peptide; 22 AA.  
 AC R44112;  
 DT 28-JUL-1994 (first entry)  
 DE HLA epitope.  
 KW Epitope; immuno-homologue; HLA; heavy chain; alpha-1 domain; allele;  
 KW cw4; superantigen; immunodiagnostic; AIDS; HIV; immunisation;  
 KW vaccines.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT region 1..9  
 FT region /label- epitope  
 FT region 14..22  
 FT region /label- epitope  
 PN WO9323427-A.  
 PD 25-NOV-1993.  
 PE 10-MAY-1993; IT0049.  
 PR 11-MAY-1992; IT-RM0350.  
 PA (SANR-) FOND CENT SAN ROMANELLO DEL MONTE TABOR.  
 PA (SUPE-) INST SUPERIORE DI SANITA.  
 PI Beretta A;  
 DR WPI; 93-386478/48.  
 PT HIV gp160 epitope(s) immunologically homologous to HLA - used to  
 PT develop agents for immuno-diagnosis and vaccines for HIV  
 PT infection  
 PS Claim 6; Page 14; 31pp; English.  
 CC The patent claims epitopes of the HIV gp160 protein which are  
 CC immunologically homologous to at least one epitope of the HLA  
 CC protein family, preferably the HLA class I protein heavy chains.  
 CC The present sequence is a region of the heavy chain alpha-1 domain,  
 CC Cw4 allele, to which the epitopes are homologous. In particular, the  
 CC epitopes are homologous to the sequences TQYKRRQAO and NLRKLRGY.  
 CC The identification of HIV gp160 epitopes homologous to HLA epitopes  
 CC can be used to develop agents for vaccines and immunodiagnosis of  
 CC HIV infection. The HIV homologous HLA regions are able to induce  
 CC anti-HLA autoantibodies and act as superantigens during the  
 CC infection course, leading to a gradual and selective loss of  
 CC certain T receptor families.  
 SQ Sequence 22 AA;



Query Match 78.4%; Score 40; DB 1; Length 22;  
Best Local Similarity 80.0%; Pred. No. 0.071;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 RVDLRTLRY 10
         ||:||| ||||
Db     12 RVNLRKLRGY 21

```

RESULT 12	
RR41206	
ID	R41206 standard; peptide; 25 AA.
AC	R41206;
DDT	15-MAR-1994 (first entry)
DE	Peptide fragment of Class I HLA peptide.
DE	Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;
KW	parasitic disease; cytotoxic T lymphocyte; modulation.
OS	Synthetic.
PN	W09317699-A.
PD	16-SEP-1993.
PPF	25-FEB-1993; U01758.
PPR	02-MAR-1992; US-844716.
PPA	(STRD ) UNIV LELAND STANFORD JUNIOR.
PPI	Clayberger CA, Krensky AM;
DR	WPI: 93-303134/38.
PT	New peptide(s) based on Class I HLA antigen domains - used for
PT	modulating cytotoxic T-lymphocyte activity towards targets
PS	Claim 9; Page 53; 61pp; English.
CCC	The peptide (or a fragment of at least 10 amino acids, joined at at
CCC	least one terminus to a sequence other than that of wild type HLA
CCC	antigen) is used to modulate cytotoxic T-lymphocyte (CTL) activity,
CCC	either by inhibition or stimulation. It can be used for
CCC	inhibiting CTL toxicity in transplantations, for inducing CTL
CCC	activity in parasitic diseases and neoplasia and in studies on viral
CCC	infection. The peptide can also be used for identifying CTLs which
CCC	bind to it and removing subsets of CTLs from a T-cell composition.
SSQ	Sequence 25 AA:

Query Match 78.4%; Score 40; DB 1; Length 25;  
Best Local Similarity 80.0%; Pred. No. 0.081;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      1  RVDLRLRGY 10
      ||| ||| |||
Db     16  RVSLRLNRGY 25

```

RESULT 13

R83072

ID R83072 standard; peptide; 25 AA.

AC R83072:

DT 16-MAY-1996 (first entry)

DE HLA-Bw46 CTL modulating peptide (Bw46.60-84).

DE Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;

KK immunosuppressant; graft versus host disorder; transplantation; therapy;

KK class I MHC; HLA-Bw46.

KW class I MHC; HLA-Bw46.

KW Synthetic.

QS QS

PN W09526979-A1.

PN PD

PD 12-OCT-1995.

PP U04349.

PP 05-APR-1995; U04349.

PP 05-APR-1994; US-222851.

PP (STRD ) UNIV LELAND STANFORD JUNIOR.

PA Clayberger C. Krensky AM, Parham P;

PI WPI: 95-358582/46.

DR Extension of acceptance period of transplants from MHC unmatched

PPT donor hosts - using Class I B75-84 MHC antigen of the recipient

PPT host

PPT Example 13; Page 32: 80pp: English.

PS R83061 R83085, R83090-R83096 and R92907-R92913 represent fragments of

CC class I major histocompatibility complex (MHC) antigens. This sequence

CC corresponds to residues 60-84 of the alpha-1 domain of the class

CC I MHC HLA-Bw46. These sequences can be used to extend the period of

CC I MHC HLA-Bw46.

R71629

ID R71629 standard; peptide; 45 AA.  
 AC R71629;  
 DT 21-SEP-1995 (first entry)  
 DE HLA-alpha-1.  
 KW HIV; gp120; HLA; beta-2-microglobulin; HLA; passive immunization  
 KW human immunodeficiency virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT peptide 16..37  
 FT /label= HLA-22  
 PN W09507465-A.  
 PD 16-MAR-1995.  
 PF 06-SEP-1994; IT0146.  
 PR 10-SEP-1993; IT-RM0613.  
 PA (SANK-) FOND CENT SAN RAFFAELE DEL MONTE TABOR.  
 PI Beretta A;  
 DR WPI; 95-123513/16.  
 PT Use of antibodies reactive with HIV and HLA molecules - for  
 PT detection of HIV infection and for providing passive immunity  
 PT against HIV infection.  
 PS Disclosure: Page 7; 26pp; English.  
 CC Antibodies able to detect both epitopes of HIV gp160 and epitopes of  
 CC beta-2-microglobulin free Class I HLA free heavy chains are pref.  
 CC prep. using gp120-C5 and HLA alpha-1 peptides (R71628-29) which  
 CC respectively include peptides HIV-26 and HLA-22.  
 SQ Sequence 45 AA;

Query Match 78.4%; Score 40; DB 1; Length 45;  
 Best Local Similarity 80.0%; Pred. No. 0.15;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRRLRGY 10  
 DB 27 RYNLRRLRGY 36

Search completed: February 8, 2000, 04:05:45  
 Job time: 9362 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 18:04:42 ; Search time 111.22 seconds  
(without alignments)  
4.241 Million cell updates/sec

Title: US-08-653-294-37  
Perfect score: 51  
Sequence: 1 RVDLRLRGY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR\_62:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	371	2 A35090	MHC class I histoc
2	46	90.2	243	2 I54459	MHC H-2K1-k - mous
3	46	90.2	344	2 S52367	RT1.A(u) protein a
4	46	90.2	344	2 I60331	RT1.A(u) alpha cha
5	46	90.2	353	2 G02922	MHC class I Lero-G
6	46	90.2	354	2 S18197	class I histocompa
7	46	90.2	359	2 I61867	MHC class I protei
8	46	90.2	361	2 B27638	MHC class I histoc
9	46	90.2	362	2 JH0538	class I histocompa
10	46	90.2	365	2 I54548	major histocompati
11	44	86.3	270	1 HLHU28	MHC class I histoc
12	44	86.3	273	1 HLHUAW	MHC class I histoc
13	44	86.3	273	1 HLHU69	MHC class I histoc
14	44	86.3	275	1 HLHU10	MHC class I histoc
15	44	86.3	332	2 S06424	MHC class I histoc
16	44	86.3	337	2 I70694	H-2D cell surface
17	44	86.3	357	2 I36965	MHC class I protei
18	44	86.3	357	2 S18198	class I histocompa
19	44	86.3	361	2 I48160	MHC class I protei
20	44	86.3	362	2 A45845	MHC class I histoc
21	44	86.3	364	2 S03535	class I histocompa
22	44	86.3	365	1 HLHUA2	MHC class I histoc
23	44	86.3	365	2 JH0534	class I histocompa
24	44	86.3	365	2 JH0536	class I histocompa
25	44	86.3	365	2 I37477	MHC class I histoc
26	44	86.3	365	2 I37482	MHC class I histoc
27	44	86.3	365	2 S01171	class I histocompa
28	44	86.3	365	2 JH0135	MHC class I histoc
29	44	86.3	365	2 A47636	MHC class I histoc
30	44	86.3	365	2 I84432	MHC class I protei

31	44	86.3	365	2 I83063	Al1.2 - human
32	44	86.3	365	2 I72171	HLA-A*33.1, HLA-A*
33	44	86.3	365	2 I72170	MHC class I histoc
34	44	86.3	365	2 I36961	MHC class I protei
35	44	86.3	365	2 I37480	HLA-A*661 antigen
36	44	86.3	365	2 I37483	HLA-A*34.2 antigen
37	44	86.3	365	2 I38443	gene HLA-A-0203 pr
38	44	86.3	365	2 I61857	MHC HLA-A2, 4a cha1
39	44	86.3	365	2 I37542	MHC class I histoc
40	44	86.3	365	2 I38442	gene HLA-A-0205 pr
41	44	86.3	365	2 I38441	gene HLA-A-6802 pr
42	44	86.3	365	2 I38519	MHC class I histoc
43	44	86.3	365	2 I56039	HLA-A*30.3 precursor
44	44	86.3	365	2 I38610	MHC class I histoc
45	44	86.3	365	2 I37470	HLA-A*0210 - human

## ALIGNMENTS

RESULT 1

A35090

MHC class I histocompatibility antigen RT1.A(a) (strain DA) - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 03-Aug-1990 #sequence\_revision 03-Aug-1990 #text\_change 23-Jul-1999

C:Accession: A35090

R:Rada, C.; Lorenzi, R.; Powis, S.J.; van den Bogaerde, J.; Parham, P.; Howard, J.C.

Proc. Natl. Acad. Sci. U.S.A. 87, 2167-2171, 1990

A:Title: Concerted evolution of class I genes in the major histocompatibility complex

A:Reference number: A35090; MUID:90192768

A:Accession: A35090

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-371 <RAD>

A:Cross-references: GB:M1018; NID:gl877415; PIDN:AA849324.1; PID:gl877416

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

F:220-285/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 51; DB 2; Length 371;  
Best Local Similarity 100.0%; Pred. No. 0.0065;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10

DB 99 RVDLRLRGY 108

RESULT 2

I54459

MHC H-2K1-k - mouse

C:Species: Mus musculus (house mouse)

C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 02-Aug-1996

C:Accession: I54459

R:Watts, S.; Davis, A.C.; Goodenow, R.S.

Immunogenetics 29, 355-357, 1989

A:Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.

A:Reference number: I54459; MUID:89233303

A:Accession: I54459

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-243 <RES>

A:Cross-references: GB:M27134; NID:gl99435; PID:g387456

C:Genetics:

A:Introns: 22/1; 112/1

Query Match 90.2%; Score 46; DB 2; Length 243;

Best Local Similarity 90.0%; Pred. No. 0.045;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10

DB 112 RVDLRLRGY 108

Db 96 RVNRLTLRGY 105

RESULT 3

S52367  
RT1.A(u) protein alpha chain - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 08-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 08-Sep-1997  
C:Accession: S52367  
R:Jolly, E.; Clarkson, C.; Howard, J.C.; Butcher, G.W.  
submitted to the EMBL Data Library, October 1994  
A:Description: Isolation of a functional cDNA encoding the RT1.A(u) MHC class I heavy chain  
A:Reference number: S52367  
A:Accession: S52367  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-344 <JOL>  
A:Cross-references: EMBL:X82106; NID:g1228090; PID:g1228091  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 90.2%; Score 46; DB 2; Length 344;  
Best Local Similarity 90.0%; Pred. No. 0.065;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLGRY 10  
||:|||||||

Db 75 RVNRLTLRGY 84

RESULT 4

I60331  
RT1.A(u) alpha chain - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 23-Feb-1997  
C:Accession: I60331  
R:Jolly, E.; Clarkson, C.; Howard, J.C.; Butcher, G.W.  
Immunogenetics 41, 326-328, 1995  
A:Title: Isolation of a functional cDNA encoding the RT1.Au MHC class I heavy chain by a  
A:Reference number: I60331; MUID:95237903  
A:Accession: I60331  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-344 <RES>  
A:Cross-references: EMBL:X82106; NID:g836643; PID:g836644  
C:Genetics:  
A:Gene: RT1.A(u)  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 90.2%; Score 46; DB 2; Length 344;  
Best Local Similarity 90.0%; Pred. No. 0.065;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLGRY 10  
||:|||||||

Db 75 RVNRLTLRGY 84

RESULT 5

G02922  
MHC class I Lero-G\*01 - golden lion tamarin (fragment)  
C:Species: Leontopithecus rosalia (golden lion tamarin)  
C:Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 23-Jul-1999  
C:Accession: G02922  
R:Cadavid, L.F.; Shuffelbotham, C.; Ruiz, F.J.; Yeager, M.; Hughes, A.L.; Watkins, D.I.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: H01971  
A:Accession: G02922  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-353 <CAD>  
A:Cross-references: EMBL:U59642; NID:g1389930; PIDN:AA897485.1; PID:g1389931

C:Genetics:  
A:Gene: Lero-G  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 90.2%; Score 46; DB 2; Length 353;  
Best Local Similarity 90.0%; Pred. No. 0.067;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLGRY 10  
||:|||||||

Db 91 RVNRLTLRGY 100

RESULT 6

S18197  
class I histocompatibility antigen Gogo-A3 alpha chain - gorilla (fragment)  
C:Species: Gorilla gorilla (gorilla)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S18197  
R:Watkins, D.I.; Chen, Z.W.; Garber, T.L.; Hughes, A.L.; Letvin, N.L.  
Immunogenetics 34, 185-191, 1991  
A:Title: Segmental exchange between MHC class I genes in a higher primate: recombinant  
A:Reference number: S18197; MUID:91372865  
A:Accession: S18197  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-354 <WAT>  
A:Cross-references: EMBL:X54375; NID:g22889; PIDN:CAA38250.1; PID:g22890  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
F:212-277/Domain: immunoglobulin homology <IMW>

Query Match 90.2%; Score 46; DB 2; Length 354;  
Best Local Similarity 90.0%; Pred. No. 0.067;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRTLGRY 10  
|||:|||||

Db 91 RVDLRTLGRY 100

RESULT 7

I61867  
MHC class I protein - orangutan (fragment)  
C:Species: Pongo pygmaeus (Orangutan)  
C:Date: 08-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 23-Jul-1999  
C:Accession: I61867  
R:Lawlor, D.A.; Warren, E.; Ward, F.E.; Parham, P.  
Immunol. Rev. 113, 147-185, 1990  
A:Title: Comparison of class I MHC alleles in humans and apes.  
A:Reference number: I36961; MUID:90201944  
A:Accession: I61867  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-359 <RES>  
A:Cross-references: GB:M30681; NID:g342845; PIDN:AAA88836.1; PID:g342846  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 90.2%; Score 46; DB 2; Length 359;  
Best Local Similarity 90.0%; Pred. No. 0.068;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLGRY 10  
||:|||||||

Db 93 RVNRLTLRGY 102

RESULT 8

B27638  
MHC class I histocompatibility antigen alpha chain precursor (BL3-7) - bovine  
C:Species: Bos primigenius taurus (cattle)

C>Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 16-Feb-1997

C:Accession: B27638

R:Ennis, P.D.; Jackson, A.P.; Parham, P.

J. Immunol. 141, 642-651, 1988

A:Title: Molecular cloning of bovine class I MHC cDNA.

A:Reference number: A92826; MUID:88258075

A:Accession: B27638

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-361 <EN>

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: heterodimer; transmembrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-361/Product: MHC class I histocompatibility antigen, BoLA alpha chain (BL3-7) #stat

F:220-285/Domain: immunoglobulin homology <IMM>

Query Match 90.2%; Score 46; DB 2; Length 361;

Best Local Similarity 90.0%; Pred. No. 0.068;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTLCY 10

DB 99 RVDLNTLCY 108

RESULT 9

JH0538

Class I histocompatibility antigen Gogo-Oko heavy chain precursor - lowland gorilla

C:Species: Gorilla gorilla gorilla (lowland gorilla)

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999

C:Accession: JH0538

R:Lawlor, D.A.; Warren, E.; Taylor, P.; Parham, P.

J. Exp. Med. 174, 1491-1509, 1991

A:Title: Gorilla class I major histocompatibility complex alleles: comparison to human a

A:Reference number: JH0534; MUID:92078860

A:Accession: JH0538

A:Molecule type: DNA

A:Residues: 1-362 <LAW>

A:Cross-references: EMBL:X60692; NID:g22908; PIDN:CAA43100.1; PID:g22909

A:Experimental source: EBV-transformed B cell

C:Genetics:

A:Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: transmembrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-362/Product: class I histocompatibility antigen heavy chain, Gogo-Oko #status predi

F:115-206/Domain: alpha-1 <AL1>

F:115-206/Domain: alpha-2 <AL2>

F:207-298/Domain: alpha-3 <AL3>

F:220-285/Domain: immunoglobulin homology <IMM>

F:299-362/Domain: intracellular #status predicted <INT>

Query Match

Best Local Similarity 90.2%; Score 46; DB 2; Length 362;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTLCY 10

DB 99 RVDLETLTCY 108

RESULT 10

I54548

major histocompatibility complex class I - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 23-Jul-1999

C:Accession: I54548

R:Walter, L.; Hiemann, C.; Heine, L.; Gunther, E.

Immunogenetics 41, 332, 1995

A:Title: Genomic organization and sequence of the rat major histocompatibility complex c

A:Reference number: I54548; MUID:95237905

A:Accession: I54548

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-365 <RES>

A:Cross-references: EMBL:X82669; NID:g642266; PIDN:CAA57986.1; PID:g1000514

C:Genetics:

A:Gene: RT1.Alu

A:Introns: 22/1; 112/1; 204/1; 296/1; 339/1; 349/1; 365/1

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match

Best Local Similarity 90.2%; Score 46; DB 2; Length 365;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRLTLCY 10

DB 96 RVDLRLTLCY 105

RESULT 11

HLH028

MHC class I histocompatibility antigen HLA-A28 alpha chain - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 02-Sep-1997

C:Accession: A02187

R:Lopez de Castro, J.A.; Strominger, J.L.; Strong, D.M.; Orr, H.T.

Proc. Natl. Acad. Sci. U.S.A. 79, 3813-3817, 1982

A:Title: Structure of crossreactive human histocompatibility antigens HLA-A28 and HLA

A:Reference number: A93919; MUID:82247941

A:Accession: A02187

A:Molecule type: protein

A:Residues: 1-270 <LOP>

C:Genetics:

A:Gene: GDB:HLA-A

A:Cross-references: GDB:119310; OMIM:142800

A:Map position: 6p21.3-6p21.3

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterodimer; membrane protein; transplantation

F:1-90/Domain: alpha-1 <EX1>

F:91-182/Domain: alpha-2 <EX2>

F:196-261/Domain: immunoglobulin homology <IMM>

F:86/Binding site: carbohydrate (Asn) #status experimental

F:101-164,203-259/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 86.3%; Score 44; DB 1; Length 270;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLTLCY 10

DB 75 RVDLRLTLCY 84

RESULT 12

HLH04W

MHC class I histocompatibility antigen HLA-A\*68 alpha chain - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 05-Sep-1997

C:Accession: A24671

R:Holmes, N.; Parham, P.

EMBO J. 4, 2849-2854, 1985

A:Title: Exon shuffling in vivo can generate novel HLA class I molecules.

A:Reference number: A91021; MUID:86055720

A:Accession: A24671

A:Molecule type: DNA

A:Residues: 1-273 <HOL>

C:Genetics:

A:Gene: GDB:HLA-A

A:Cross-references: GDB:119310; OMIM:142800

A:Map position: 6p21.3-6p21.3

A:Introns: 89/3

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; heterodimer; transmembrane protein; transplantation antigen  
F:195-260/Domain: immunoglobulin homology <IMM>  
F:85/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.3%; Score 44; DB 1; Length 273;  
Best Local Similarity 90.0%; Pred. No. 0.13;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RVDLRTLRGY 10  
|||||  
Db 74 RVDLGLTLRGY 83

## RESULT 13

HLHU69  
MHC class I histocompatibility antigen HLA-Aw69 alpha chain - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 05-Sep-1997  
C:Accession: B24671  
R:Holmes, N.; Parham, P.  
EMBO J. 4, 2849-2854, 1985  
A:Title: Exon shuffling in vivo can generate novel HLA class I molecules.  
A:Reference number: A91021; MUID:86055720  
A:Accession: B24671  
A:Molecule type: DNA  
A:Residues: 1-273 <HOL>  
C:Genetics:

A:Gene: GDB:HLA-A  
A:Cross-references: GDB:119310; OMIM:142800  
A:Map position: 6p21.3-6p21.3  
A:Introns: 89/3

C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; heterodimer; transmembrane protein; transplantation antigen  
F:195-260/Domain: immunoglobulin homology <IMM>  
F:85/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.3%; Score 44; DB 1; Length 273;  
Best Local Similarity 90.0%; Pred. No. 0.13;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RVDLRTLRGY 10  
|||||  
Db 74 RVDLGLTLRGY 83

## RESULT 14

HLHU10

MHC class I histocompatibility antigen HLA-A10 alpha chain - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 22-Jun-1999  
C:Accession: B24512  
R:Davidson, W.F.; Kress, M.; Khoury, G.; Jay, G.  
J. Biol. Chem. 260, 13414-13423, 1985  
A:Title: Comparison of HLA class I gene sequences. Derivation of locus-specific oligonuc  
A:Reference number: A92500; MUID:86033791  
A:Accession: B24512  
A:Molecule type: DNA  
A:Residues: 1-275 <DAV>  
A:Cross-references: GB:M11887; NID:g184157; PIDN:AAA52656.1; PID:g184158

C:Genetics:  
A:Gene: GDB:HLA-A  
A:Cross-references: GDB:119310; OMIM:142800  
A:Map position: 6p21.3-6p21.3  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; heterodimer; transmembrane protein; transplantation antigen  
F:130-195/Domain: immunoglobulin homology <IMM>  
F:20/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.3%; Score 44; DB 1; Length 275;

Best Local Similarity 90.0%; Pred. No. 0.13;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RVDLRTLRGY 10  
|||||  
Db 9 RVDLGLTLRGY 18

## RESULT 15

S06424

MHC class I histocompatibility antigen Ch25 alpha chain precursor - chimpanzee  
N:Alternate names: MHC Ch1A chain  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999  
C:Accession: S06424; I36959  
R:Lawlor, D.A.; Ward, F.E.; Ennis, P.D.; Jackson, A.P.; Parham, P.  
Nature 335, 268-271, 1988  
A:Title: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees.  
A:Reference number: S06424; MUID:88319000

A:Accession: S06424  
A:Molecule type: mRNA  
A:Residues: 1-332 <LAW>

R:Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.  
J. Immunol. 142, 3937-3950, 1989

A:Title: Diversity and diversification of HLA-A,B,C alleles.  
A:Reference number: I36956; MUID:89235215

A:Accession: I36959

A:Molecule type: mRNA

A:Residues: 1-332 <RES>

A:Cross-references: GB:M24047; NID:g176818; PIDN:AAA35426.1; PID:g553155  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein; membrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-114/Domain: alpha-1 #status predicted <EX1>  
F:115-206/Domain: alpha-2 #status predicted <EX2>  
F:220-285/Domain: immunoglobulin homology <IMM>  
F:307-331/Domain: transmembrane #status predicted <TM>  
F:110/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:125-188,227-283/Disulfide bonds: #status predicted

Query Match 86.3%; Score 44; DB 2; Length 332;  
Best Local Similarity 90.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RVDLRTLRGY 10  
|||||  
Db 99 RVDLGLTLRGY 108

Search completed: February 7, 2000, 18:04:42  
Job time: 22208 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:26:02 ; Search time 68.63 seconds  
(without alignments)  
4.352 Million cell updates/sec

Title: US-08-653-294-37  
Perfect score: 51  
Sequence: 1 RVDLRLRGY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	371	1 HA12_RAT	P16391 rattus norv
2	46	90.2	358	1 HLA_E_HUMAN	P13747 homo sapien
3	46	90.2	359	1 HLA_E_PONPY	P16212 pongo pygma
4	46	90.2	362	1 OKO_GORGO	P30388 gorilla gor
5	46	90.2	364	1 HA1B_BOVIN	P13753 bos taurus
6	44	86.3	273	1 HA69_HUMAN	P10316 homo sapien
7	44	86.3	275	1 IXX_HUMAN	P10313 homo sapien
8	44	86.3	362	1 HA19_CANFA	P18466 canis famil
9	44	86.3	365	1 HA01_GORGO	P30375 gorilla gor
10	44	86.3	365	1 HA01_PANTR	P16209 pan troglod
11	44	86.3	365	1 HA01_PONPY	P16211 pongo pygma
12	44	86.3	365	1 HA02_HUMAN	P01892 homo sapien
13	44	86.3	365	1 HA03_GORGO	P30377 gorilla gor
14	44	86.3	365	1 HA03_PANTR	P13748 pan troglod
15	44	86.3	365	1 HA04_PANTR	P13749 pan troglod
16	44	86.3	365	1 HA11_HUMAN	P13746 homo sapien
17	44	86.3	365	1 HA30_HUMAN	P16188 homo sapien
18	44	86.3	365	1 HA31_HUMAN	P16189 homo sapien
19	44	86.3	365	1 HA33_HUMAN	P16190 homo sapien
20	44	86.3	365	1 HA34_HUMAN	P30453 homo sapien
21	44	86.3	365	1 HA66_HUMAN	P30457 homo sapien
22	44	86.3	365	1 HA68_HUMAN	P01891 homo sapien
23	44	86.3	365	1 HA74_HUMAN	P30459 homo sapien
24	44	86.3	369	1 HA1B_MOUSE	P01901 mus musculu
25	44	86.3	370	1 HA03_HUMAN	P04439 homo sapien
26	41	80.4	360	1 HA1A_BOVIN	P13752 bos taurus
27	40	78.4	342	1 ICXX_HUMAN	P10321 homo sapien
28	40	78.4	362	1 IB44_HUMAN	P30484 homo sapien
29	40	78.4	365	1 HA01_SAGOE	P30515 saguinus oe
30	40	78.4	365	1 IC01_HUMAN	P30499 homo sapien
31	40	78.4	366	1 IC01_PANTR	P30686 pan troglod
32	40	78.4	366	1 IC02_GORGO	P30385 gorilla gor
33	40	78.4	366	1 IC02_HUMAN	P30500 homo sapien
34	40	78.4	366	1 IC03_GORGO	P30386 gorilla gor

```

35 40 78.4 366 1 IC03_HUMAN P30501 homo sapien
36 40 78.4 366 1 IC04_GORGO P30387 gorilla gor
37 40 78.4 366 1 IC04_HUMAN P30502 homo sapien
38 40 78.4 366 1 IC05_HUMAN P04222 homo sapien
39 40 78.4 366 1 IC06_HUMAN P30503 homo sapien
40 40 78.4 366 1 IC07_HUMAN P30504 homo sapien
41 40 78.4 366 1 IC11_HUMAN Q29631 homo sapien
42 40 78.4 366 1 IC12_HUMAN P30505 homo sapien
43 40 78.4 366 1 IC13_HUMAN P30506 homo sapien
44 40 78.4 366 1 IC14_HUMAN P30507 homo sapien
45 40 78.4 366 1 IC15_HUMAN P30508 homo sapien

```

## ALIGNMENTS

RESULT 1

```

HA12_RAT
ID HA12_RAT STANDARD; PRT; 371 AA.
AC P16391; 002940;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE RT1 CLASS I HISTOCOMPATIBILITY ANTIGEN, AA ALPHA CHAIN PRECURSOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90192768.
RA RADA C., LORENZI R., POWIS S.J., DEN BOGAERDE J., PARHAM P.,
RA HOWARD J.C.
RT "Concerted evolution of class I genes in the major histocompatibility
complex of murine rodents."
RL Proc. Natl. Acad. Sci. U.S.A. 87:2167-2171(1990).
RN [2]
RP REVISIONS.
RC STRAIN-DA; TISSUE=SPLEEN, AND LYMPH NODE;
RA HOWARD J.C.
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31018; AAB49324.1; -
CC HSSP; P01901; 2CKB.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00126; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 371
FT RT1 CLASS I HISTOCOMPATIBILITY ANTIGEN,
AA ALPHA CHAIN.
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 311
FT TRANSMEM 312 336
FT DOMAIN 337 371
FT CARBOHYD 110 110
FT CARBOHYD 280 280
FT SEQUENCE 371 AA; 41830 MW; 4E8C6EFF CRC32;
SQ

```

FT	DOMAIN	296	305
FT	TRANSMEM	306	329
FT	DOMAIN	330	358
FT	DISULFID	122	185
FT	DISULFID	224	280
FT	CARBOHYD	107	107
FT	VARIANT	10	10
FT	VARIANT	104	104
FT	SEQUENCE	358 AA;	40130 MW; 3D79F233 CRC32;

CONNECTING PEPTIDE.

CYTOPLASMIC TAIL.

BY SIMILARITY.  
S -> L (IN E\*0102).  
/FTID=VAR\_004394.  
G->R (IN E\*0102).  
/FTID=VAR\_004395.

Query Match	90.2%;	Score 46;	DB 1;	Length 358;
Best Local Similarity	90.0%;	Pred. No. 0.023;		
Matches 9:	Conservative	1;	Mismatches 0;	Indels 0;
Gaps	0;			

```

      104      104      0.000000
      FT      SEQUENCE      358 AA;      40130 MW;      3D79F233 CRC32;
      /FTId=VAR_004395;
      SQ      QUERY      1 RVDLRLRGY 10
      Query Match      90.2%;      Score 46;      DB 1;      Length 358;
      Best Local Similarity 90.0%;      Pred. No. 0.023;
      Matches 9;      Conservative 1;      Mismatches 0;      Indels 0;      Gaps 0;
      Db      96 RVDLRLRGY 105
      ||:|||||
      ||:|||||
  
```

RESULT	3			
HLAE_PONPY		STANDARD;	PRT;	359 AA.
ID	HLAE_PONPY			
AC	P16212;			
DT	01-APR-1990	(Rel. 14, Created)		
DT	01-APR-1990	(Rel. 14, Last sequence update)		
DT	15-JUL-1999	(Rel. 38, Last annotation update)		
DE	CLASS I HISTOCOMPATIBILITY ANTIGEN, E-1 ALPHA CHAIN PRECURSOR			
DE	(FRAGMENT)			

DI 15 000 1399; (Acc. 507; East annotation update)  
 CLASS 1 HISTOCOMPATIBILITY ANTIGEN, E-1 ALPHA CHAIN PRECURSOR  
 (FRAGMENT).  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Pongo.  
 RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE: 90201944.  
RA LAWLER D.A., WARREN E., WARD F.E., PARHAM P.;  
RA "Comparison of class I MHC alleles in humans and apes.";  
RT Immunol. Rev. 113:147-185(1990).  
RL -I- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -I- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M30681; AAA89836.1; -

CC DR HSSP; P03989; 1HSA.

CC DR PROSITE; PS00290; IG\_MHC; 1.

CC DR

[illegible]



Query Match 90.2%; Score 46; DB 1; Length 359;  
 Best Local Similarity 90.0%; Pred. No. 0.023; Mismatches 1; Indels 0; Gaps 0;  
 Matches 9; Conservative 0;

QY 1 RVDLRTLGRY 10  
 ||:|||||  
 DB 93 RVDLRTLGRY 102

RESULT 4  
 ID LOKO\_GORGO STANDARD; PRT; 362 AA.  
 AC P30388;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-OKO ALPHA CHAIN PRECURSOR.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92078860.  
 RA LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;  
 RT "Gorilla Class I major histocompatibility complex alleles: comparison to human and chimpanzee class I.";  
 RL J. Exp. Med. 174:1491-1509(1991).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X60692; CAA43100.1; -  
 DR PIR; JH0538; JH0538.  
 DR HSP; P30460; IAGB.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; 19; 1.  
 DR MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 308  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 362  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 FT CARBOHYD 110 110  
 SQ SEQUENCE 362 AA; 40755 MW; CD795172 CRC32;

Query Match 90.2%; Score 46; DB 1; Length 362;  
 Best Local Similarity 90.0%; Pred. No. 0.023; Mismatches 0; Indels 1; Gaps 0;  
 Matches 9; Conservative 0;

QY 1 RVDLRTLGRY 10  
 |||||  
 DB 99 RVDLRTLGRY 108

RESULT 5  
 ID HALB\_BOVIN STANDARD; PRT; 364 AA.  
 AC P3753;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-JAN-1990 (Rel. 13, Last annotation update)  
 DE BOLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN BL3-7 PRECURSOR.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 88258075.  
 RA ENNIS P.D., JACKSON A.P., PARHAM P.;  
 RT "Molecular cloning of bovine class I MHC CDNA.";  
 RL J. Immunol. 141:642-651(1988).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M21043; AAA30641.1; -  
 DR PIR; B27638; B27638.  
 DR HSP; P30460; IAGB.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFAM; PF00047; 19; 1.  
 DR MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 364  
 FT DOMAIN 28 117  
 FT DOMAIN 118 209  
 FT DOMAIN 210 301  
 FT DOMAIN 302 310  
 FT TRANSMEM 311 331  
 FT DOMAIN 332 364  
 FT SIMILAR 210 301  
 FT CARBOHYD 106 106  
 FT CARBOHYD 113 113  
 FT DISULFID 128 191  
 FT DISULFID 230 286  
 SQ SEQUENCE 364 AA; 41513 MW; A27F581C CRC32;

Query Match 90.2%; Score 46; DB 1; Length 364;  
 Best Local Similarity 90.0%; Pred. No. 0.023; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 0;

QY 1 RVDLRTLGRY 10  
 |||||  
 DB 102 RVDLRTLGRY 111

RESULT 6  
 ID JA69\_HUMAN STANDARD; PRT; 273 AA.  
 AC P10316;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, AW-69(A-28) ALPHA CHAIN

```

DE (FRAGMENT).
GN HLA-A OR HLA.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (A*6901).
RX MEDLINE: 86055720.
RA HOLMES N., PARHAM P.;
RT "Exon shuffling in vivo can generate novel HLA class I molecules.";
RL EMBO J. 4:2849-2854(1985).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -!- POLYMORPHISM: THE ONLY ALLELE OF AW-69 KNOWN IS A*6901 WHICH IS
CC SHOWN HERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03158; ; NOT_ANNOTATED_CDS.
DR EMBL: X03159; ; NOT_ANNOTATED_CDS.
DR FIR: B24671; HLH069.
DR HSP: P01892; LA0D.
DR MIM: 142800;
DR PFAM: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein.
KW NON_TER 1
FT DOMAIN <1 89 EXTRACELLULAR ALPHA-1.
FT DOMAIN 90 180 EXTRACELLULAR ALPHA-2.
FT DOMAIN 181 273 EXTRACELLULAR ALPHA-3.
FT CARBOHYD 85 85 BY SIMILARITY.
FT DISULFID 100 163 BY SIMILARITY.
FT DISULFID 202 258 BY SIMILARITY.
FT NON_TER 273 273
SQ SEQUENCE 273 AA; 31677 MW; EEBFB366 CRC32;

```

```

Query Match 86.3%; Score 44; DB 1; Length 273;
Best Local Similarity 90.0%; Pred. No. 0.044;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRTLCY 10
DB 74 RVDLGTLCY 83

```

```

RESULT 7
ID LAXX_HUMAN STANDARD; PRT; 275 AA.
AC P10313;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-10 ALPHA CHAIN (FRAGMENT).
GN HLA-A OR HLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86033791.
RA DAVIDSON W.F., KRESS M., KHOURY G., JAY G.;
RT "Comparison of HLA class I gene sequences. Derivation of
RT locus-specific oligonucleotide probes specific for HLA-A, HLA-B, and

```

```

RT HLA-C genes.";
RL J. Biol. Chem. 260:13414-13423(1985).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M11887; AAA52656.1; -.
DR FIR: B24512; HLH010.
DR HSP: P01892; 1HHH.
DR MIM: 142800;
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00129; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein.
KW NON_TER 1
FT DOMAIN <1 24 EXTRACELLULAR ALPHA-1.
FT DOMAIN 25 116 EXTRACELLULAR ALPHA-2.
FT DOMAIN 117 208 EXTRACELLULAR ALPHA-3.
FT DOMAIN 209 218 CONNECTING PEPTIDE.
FT TRANSMEM 219 242
FT DOMAIN 243 275 CYTOPLASMIC TAIL.
FT CARBOHYD 20 20 BY SIMILARITY.
FT DISULFID 35 98 BY SIMILARITY.
FT DISULFID 137 193 BY SIMILARITY.
FT SEQUENCE 275 AA; 30548 MW; 8B232F3C CRC32;

Query Match 86.3%; Score 44; DB 1; Length 275;
Best Local Similarity 90.0%; Pred. No. 0.045;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRTLCY 10
DB 9 RVDLGTLCY 18

```

```

RESULT 8
ID HAL9_CANFA STANDARD; PRT; 362 AA.
AC P18466;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A9/A9 ALPHA CHAIN PRECURSOR.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90316611.
RA SARMIENTO U.M., STORB R.;
RT "Nucleotide sequence of a dog class I cDNA clone.";
RL Immunogenetics 31:400-404(1990).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

FT	DOMAIN	25	114	GOGO-A0101 ALPHA CHAIN.
FT	DOMAIN	115	206	EXTRACELLULAR ALPHA-1.
FT	DOMAIN	207	298	EXTRACELLULAR ALPHA-2.
FT	DOMAIN	299	308	EXTRACELLULAR ALPHA-3.
FT	DOMAIN	309	332	CONNECTING PEPTIDE.
FT	TRANSMEM	333	365	CYTOPLASMIC TAIL.
FT	DOMAIN	333	365	BY SIMILARITY.
FT	DISULFID	125	188	BY SIMILARITY.
FT	DISULFID	227	283	BY SIMILARITY.
FT	CARBOHYD	110	110	BY SIMILARITY.
SQ	SEQUENCE	365 AA;	40829 MW;	406826C4 CRC32;

Query Match 86.3%; Score 44; DB 1; Length 365;  
Best Local Similarity 90.0%; Pred. No. 0.062;  
Matches 9; Conservative 0; Mismatches 1; Indels

QY	1	RVDLRTLRGY	10
Db	99	RVDLGTLRGY	108

RESULT	10
IAOL_PANTR	
ID	IAOL_PANTR
AC	P16309;
DT	01-APR-1990 (Rel. 14, Created)
DT	01-APR-1990 (Rel. 14, Last sequence update)
DT	01-APR-1993 (Rel. 25, Last annotation update)
DE	CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-2 ALPHA CHAIN PRECURSOR.
OS	Pan troglodytes (Chimpanzee).

CC -I- FUNCTION INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.

-1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M30678; AAA87970.1; -.

DR HSP; P01892; 1HH.  
DR PROSITE; P800290; IG\_MHC; 1.  
DR PFAM; PF00047; ig; 1.  
DR PFAM; PF00129; MHC\_I; 1.  
DR MHC I; Transmembrane; Glycoprotein; Signal.  
KW SIGNAL 1 24

CHAIN	25	365	CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-2 ALPHA CHAIN.
FT			
FT			
FT	25	114	EXTRACELLULAR ALPHA-1.
FT	115	206	EXTRACELLULAR ALPHA-2.
FT	207	298	EXTRACELLULAR ALPHA-3.
FT	299	308	CONNECTING PEPTIDE.
FT	309	332	
FT	333	365	CYTOPLASMIC TAIL.
FT	125	188	BY SIMILARITY.
FT	227	283	BY SIMILARITY.
FT	110	110	BY SIMILARITY.
SQ	365 AA;	440848 MW;	56C5FF0E CRC32;

DT DT DT

GNOS OC

RP RX RA RT RT RL RN RP RX RA RA RT RT RL RN RP RX RA RT RT RL

RN RP RX RA RA RT RL RN RP RX RA RT RT RL RN RP RX RA RA RT RT RT RL RN RP RX RA RT

RL RL RN RP RX RA RA RT RT RL RN RP RX RA

RT "Multiple genetic mechanisms have contributed to the generation of  
RT the HLA-A2/A28 family of class I MHC molecules.";  
RL J. Immunol. 139:936-941(1987).  
RN [10]  
RP SEQUENCE FROM N.A. (A\*0203/A\*0205).  
RA DOMENA J.D.;  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
RN [11]  
RP SEQUENCE OF 9-365 FROM N.A. (A\*0204).  
RX MEDLINE: 92039809.  
RA CASTANO A.R., LOPEZ DE CASTRO J.A.;  
RT "Structure of the HLA-A\*0204 antigen, found in South American  
RT Indians. Spatial clustering of HLA-A2 subtype polymorphism.";  
RL Immunogenetics 34:281-285(1991).  
RN [12]  
RP SEQUENCE OF 9-365 FROM N.A. (A\*0204).  
RX MEDLINE: 92269956.  
RA WAYKINS D.I., MCADAM S.N., LIU X., STANG C.R., MILFORD E.L.,  
RA LEVINE C.G., GARBER T.L., DOGON A.L., LORD C.I., GHIM S.H.,  
RA TROUP G.M., HUGHES A.L., LETVIN N.L.;  
RT "New recombinant HLA-B alleles in a tribe of South American  
RT Amerindians indicate rapid evolution of MHC class I loci.";  
RL Nature 357:329-333(1992).  
RN [13]  
RP SEQUENCE FROM N.A. (A\*0206).  
RX MEDLINE: 92235215.  
RA PARHAM P., LAWLER D.A., LOWEN C.E., ENNIS P.D.;  
RT "Diversity and diversification of HLA-A,B,C alleles.";  
RL J. Immunol. 142:3937-3950(1989).  
RN [14]  
RP PARTIAL SEQUENCE (A\*0206).  
RX MEDLINE: 86305811.  
RA EZQUERRA A., DOMENECH N., VAN DER POEL J., STROMINGER J.L., VEGA M.A.,  
RA LOPEZ DE CASTRO J.A.;  
RT "Molecular analysis of an HLA-A2 functional variant CIA defined by  
RT cytolytic T lymphocytes.";  
RL J. Immunol. 137:1642-1649(1986).  
RN [15]  
RP PARTIAL SEQUENCE (A\*0207).  
RX MEDLINE: 88113844.  
RA DOMENECH N., EZQUERRA A., CASTANO R., LOPEZ DE CASTRO J.A.;  
RT "Structural analysis of HLA-A2.4 functional variant KNE. Implications  
RT for the mapping of HLA-A2-specific T-cell epitopes.";  
RL Immunogenetics 27:196-202(1988).  
RN [16]  
RP PARTIAL SEQUENCE (A\*0208).  
RX MEDLINE: 88314183.  
RA DOMENECH N., CASTANO R., GOULMY E., LOPEZ DE CASTRO J.A.;  
RT "Molecular analysis of HLA-A2.4 functional variant KLO: close  
RT structural and evolutionary relatedness to the HLA-A2.2 subtype.";  
RL Immunogenetics 28:143-152(1988).  
RN [17]  
RP PARTIAL SEQUENCE (A\*0209).  
RX MEDLINE: 88185100.  
RA CASTANO R., EZQUERRA A., DOMENECH N., LOPEZ DE CASTRO J.A.;  
RT "An HLA-A2 population variant with structural polymorphism in the  
RT alpha 3 region.";  
RL Immunogenetics 27:345-355(1988).  
RN [18]  
RP SEQUENCE FROM N.A. (A\*0210).  
RX MEDLINE: 89122133.  
RA EPSTEIN H., KENNEDY L., HOLMES N.;  
RT "An Oriental HLA-A2 subtype is closely related to a subset of  
RT caucasoid HLA-A2 alleles.";  
RL Immunogenetics 29:112-116(1989).  
RN [19]  
RP SEQUENCE OF 9-365 FROM N.A. (A\*0211).  
RX MEDLINE: 92218010.  
RA CASTANO A.R., LOPEZ DE CASTRO J.A.;  
RT "Structure of the HLA-A\*0211 (A2.5) subtype: further evidence for  
RT selection-driven diversification of HLA-A2 antigens.";  
RL Immunogenetics 35:344-346(1992).  
RN [20]

RP SEQUENCE FROM N.A. (A\*0213).  
RX MEDLINE: 94222455.  
RA BARBER D.F., FERNANDEZ J.M., LOPEZ DE CASTRO J.A.;  
RT "Primary structure of a new HLA-A2 subtype: HLA-A\*0213.";  
RL Immunogenetics 39:378-378(1994).  
RN [21]  
RP SEQUENCE FROM N.A. (A\*0216).  
RX MEDLINE: 95278976.  
RA BAROUCH D., KRAUSA P., BODMER J., BROWNING M.J., MCMICHAEL A.J.;  
RT "Identification of a novel HLA-A2 subtype, HLA-A\*0216.";  
RL Immunogenetics 41:388-388(1995).  
RN [22]  
RP SEQUENCE FROM N.A. (A\*0217).  
RX MEDLINE: 95381236.  
RA SELVAKUMAR A., GRANJA C.B., SALAZAR M., ALOSCO S.M., YUNIS E.J.,  
RA DUPONT B.;  
RT "A novel subtype of A2 (A\*0217) isolated from the South American  
RT Indian B-cell line AMALA.";  
RL Tissue Antigens 45:343-347(1995).  
RN [23]  
RP SEQUENCE FROM N.A. (A\*0218).  
RX MEDLINE: 95381236.  
RA KASHIWASE K., TOKUNAGA K., ISHIKAWA Y., OOHASHI H., HASHIMOTO M.,  
RA AKAZA T., TADOKORO K., JUJI T.;  
RT "A new A2 sequence HLA-A2K from Japanese.";  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [24]  
RP SEQUENCE FROM N.A. (A\*0220).  
RX MEDLINE: 97161038.  
RA FLEISCHHAUER K., ZINO E., MAZZI B., SEVERINI G.M., BENAZZI E.,  
RA BORDIGNON C.;  
RT "HLA-A\*02 subtype distribution in Caucasians from northern Italy:  
RT Identification of A\*0220.";  
RL Tissue Antigens 48:673-679(1996).  
RN [25]  
RP SEQUENCE FROM N.A. (A\*0221).  
RX MEDLINE: 97161038.  
RA SZMANIA S., BAXTER-LOWE L.A.;  
RT "Nucleotide sequence of a novel HLA-A2 gene.";  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [26]  
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF A\*0201.  
RX MEDLINE: 88014204.  
RA BJORKMAN P.J., SAPER M.A., SAMRAOUI B., BENNETT W.S.,  
RA STROMINGER J.L., WILEY D.C.;  
RT "Structure of the human class I histocompatibility antigen, HLA-A2.";  
RL Nature 329:506-512(1987).  
RN [27]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF A\*0201.  
RX MEDLINE: 91245570.

Query Match 86.3%; Score 44; DB 1; Length 365;  
Best Local Similarity 90.0%; Pred. No. 0.062;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 RVDLRLTRY 10  
DB 99 RVDLGLTRY 108

RESULT 13

ID 1A03\_GORGO STANDARD; PRT; 365 AA.

AC P30377;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-A0401 ALPHA CHAIN PRECURSOR.

OS Gorilla gorilla gorilla (Lowland gorilla).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominidae; Gorilla.



DR EMBL; X13114; CAA31506.1; -  
 DR PIR; S01171; S01171.  
 DR HSP; P01891; 2HLA.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PFW; PF00047; 1g; 1.  
 DR PFW; PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 365  
 FT CHAIN 25 365  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 308  
 FT TRANSMEM 309 332  
 FT DOMAIN 333 365  
 FT DOMAIN 125 188  
 FT DISULFID 227 283  
 FT CARBOHYD 110 110  
 SQ SEQUENCE 365 AA; 40656 MW; 94G6AF8 CRC32;  
 CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 A-126 ALPHA CHAIN.  
 EXTRACELLULAR ALPHA-1.  
 EXTRACELLULAR ALPHA-2.  
 EXTRACELLULAR ALPHA-3.  
 CONNECTING PEPTIDE.  
 CYTOPLASMIC TAIL.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.

Query Match 86.3%; Score 44; DB 1; Length 365;  
 Best Local Similarity 90.0%; Pred. No. 0.062;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRTLCY 10  
 Db 99 RVDLGTLCY 108  
 |||||

Search completed: February 8, 2000, 01:26:03  
 Job time: 1563 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 19:16:19 ; Search time 176.54 Seconds  
(without alignments)  
3.927 Million cell updates/sec

Title: US-08-653-294-37  
Perfect score: 51  
Sequence: 1 RVDLRLRGY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SPTREMBL\_12.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	344	7 002949	O02949 rattus norv
2	51	100.0	345	7 095567	O95567 rattus norv
3	51	100.0	345	7 P79599	P79599 rattus norv
4	51	100.0	345	11 P70527	P70527 rattus norv
5	51	100.0	346	7 002953	O02953 rattus norv
6	51	100.0	356	7 Q30441	Q30441 callithrix
7	51	100.0	367	7 Q31254	Q31254 rattus norv
8	48	94.1	364	7 Q19243	Q19243 sus scrofa
9	47	92.2	356	7 Q30221	Q30221 ateles belz
10	46	90.2	90	7 Q19744	Q19744 homo sapien
11	46	90.2	144	7 Q62875	Q62875 pithecia pi
12	46	90.2	166	7 Q46720	Q46720 bos taurus
13	46	90.2	166	7 Q46722	Q46722 bos taurus
14	46	90.2	166	7 Q46734	Q46734 bos taurus
15	46	90.2	166	7 Q46743	Q46743 bos taurus
16	46	90.2	167	7 Q46726	Q46726 bos taurus
17	46	90.2	167	7 Q46731	Q46731 bos taurus
18	46	90.2	167	7 Q46737	Q46737 bos taurus
19	46	90.2	167	7 Q46747	Q46747 bos taurus
20	46	90.2	167	7 Q46748	Q46748 bos taurus

21	46	90.2	167	7 046746	O46746 bos taurus
22	46	90.2	167	7 046749	O46749 bos taurus
23	46	90.2	167	7 046752	O46752 bos taurus
24	46	90.2	181	7 Q29896	Q29896 homo sapien
25	46	90.2	336	7 Q30287	Q30287 bos taurus
26	46	90.2	343	7 Q78030	Q78030 bos taurus
27	46	90.2	344	7 Q95571	Q95571 rattus norv
28	46	90.2	345	7 Q95565	Q95565 rattus norv
29	46	90.2	353	7 Q30537	Q30537 leontopithe
30	46	90.2	353	7 Q30538	Q30538 leontopithe
31	46	90.2	354	7 Q78204	Q78204 gorilla gor
32	46	90.2	356	7 Q30445	Q30445 callithrix
33	46	90.2	357	7 Q30485	Q30485 equus cabal
34	46	90.2	357	7 Q30894	Q30894 pithecia pi
35	46	90.2	365	7 Q95577	Q95577 rattus norv
36	46	90.2	365	7 Q31256	Q31256 rattus norv
37	45	88.2	206	7 Q43906	Q43906 homo sapien
38	45	88.2	337	7 Q62934	Q62934 rattus norv
39	44	86.3	89	7 Q19558	Q19558 homo sapien
40	44	86.3	89	7 Q78085	Q78085 homo sapien
41	44	86.3	89	7 Q98051	Q98051 canis famil
42	44	86.3	89	7 Q98052	Q98052 canis famil
43	44	86.3	89	7 Q98053	Q98053 canis famil
44	44	86.3	89	7 Q98054	Q98054 canis famil
45	44	86.3	89	7 Q98055	Q98055 canis famil

## ALIGNMENTS

```

RESULT 1
O02949          PRELIMINARY;      PRT;    344 AA.
AC  O02949:
DT  01-JUL-1997 (TREMBlrel. 04, Created)
DT  01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE  01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE  MATURE MHC CLASS I ALPHA CHAIN (FRAGMENT).
GN  RT1.A2F.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-LEWIS.1F; TISSUE-SPENOCYTES;
RA  JOLY E., LE ROULE A.F., GONZALEZ A.L., MEHLING B., STEVENS J.,
RA  CORDWELL W.J., HUENIG T., HOWARD J.C., BUTCHER G.W.;
RL  Curr. Biol. 0:0-0(0).
DR  EMBL; Y13579; CAA73907.1; -.
DR  HSSP; P01901; 1KRG.
DR  PFAM; PF00047; Ig; 1.
DR  PFAM; PF00129; MHC_I; 1.
KW  MHC.
FT  NON_TER 1
FT  NON_TER 344
SQ  SEQUENCE 344 AA; 39117 MW; 7AB3EF04 CRC32;

Query Match      100.0%; Score 51; DB 7; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 RVDLRLRGY 10
    |||||
Db  75 RVDLRLRGY 84

RESULT 2
O95567          PRELIMINARY;      PRT;    345 AA.
AC  O95567:
DT  01-FEB-1997 (TREMBlrel. 02, Created)
DT  01-FEB-1997 (TREMBlrel. 02, Last sequence update)

```

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE MATURE ALPHA CHAIN OF MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I  
 DE ANTIGEN (FRAGMENT).  
 GN RT1.A2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RT1(0); TISSUE-SPLEEN;  
 RA JOLY E., LE ROLLE A.F., GONZALEZ A.L., MEHLING B., STEVENS J.,  
 RA COADWELL W.J., HUENIG T., HOWARD J.C., BUTCHER G.W.;  
 RL Curr. Biol. 0:0-0(0).  
 DR EMBL: X90372; CAA62022.1; -  
 DR HSSP: P01901; 1KBG.  
 DR PFAM: PF00047; 1g; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 345 AA; 39139 MW; 8E486EB4 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 0.022;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLRY 10  
 |||||  
 DB 75 RVDLRTLRY 84

RESULT 3  
 ID P79599 PRELIMINARY; PRT; 345 AA.  
 AC P79599; Q95564;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE MATURE ALPHA CHAIN OF MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I  
 DE ANTIGEN (FRAGMENT).  
 GN RT1.A1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RT1(0); TISSUE-SPLEEN;  
 RA JOLY E., LE ROLLE A.F., GONZALEZ A.L., MEHLING B., STEVENS J.,  
 RA COADWELL W.J., HUENIG T., HOWARD J.C., BUTCHER G.W.;  
 RL Curr. Biol. 0:0-0(1999).  
 DR EMBL: X90373; CAA62023.1; -  
 DR HSSP: P01900; 1BII.  
 DR PFAM: PF00047; 1g; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 345 AA; 39221 MW; 602EBF79 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 0.022;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLRY 10  
 |||||  
 DB 75 RVDLRTLRY 84

RESULT 4  
 ID P70527 PRELIMINARY; PRT; 345 AA.  
 AC P70527  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE PROTEIN INVOLVED IN PEPTIDE PRESENTATION (FRAGMENT).  
 GN RT1.AG.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NEDH; TISSUE-SPLENOCYTES;  
 RA JOLY E., LE ROLLE A.F., GONZALEZ A.L., MEHLING B., STEVENS J.,  
 RA COADWELL W.J., HUENIG T., HOWARD J.C., BUTCHER G.W.;  
 RL Curr. Biol. 0:0-0(0).  
 DR EMBL: Y08532; CAA69848.1; -  
 DR HSSP: P01900; 1BII.  
 DR PFAM: PF00047; 1g; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 345 AA; 38953 MW; 5A4ED67C CRC32;

Query Match 100.0%; Score 51; DB 11; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 0.022;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLRY 10  
 |||||  
 DB 75 RVDLRTLRY 84

RESULT 5  
 ID O02953 PRELIMINARY; PRT; 346 AA.  
 AC O02953;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE MHC CLASS I ALPHA CHAIN (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LEW 1F (F HAPLOTYPE);  
 RA JOLY E., LE ROLLE A.F., GONZALEZ A.L., MEHLING B., STEVENS J.,  
 RA COADWELL W.J., HUENIG T., HOWARD J.C., BUTCHER G.W.;  
 RL Curr. Biol. 0:0-0(0).  
 DR EMBL: Y14014; CAA74333.1; -  
 DR HSSP: P01901; 1KEG.  
 DR PFAM: PF00047; 1g; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 346 AA; 39103 MW; FCAA4D64 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 0.023;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVDLRTLRY 10  
 |||||  
 DB 75 RVDLRTLRY 84

RESULT 6  
 ID Q30441 PRELIMINARY; PRT; 356 AA.  
 AC Q30441  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE MHC CLASS I CAJA-G\*01 (FRAGMENT).

OS Callithrix jacchus (Common marmoset).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98070787.  
 RA CADAVID L.F., SHUFFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.,  
 RA WATKINS D.I.;  
 RT "Evolutionary instability of the major histocompatibility complex  
 class I loci in New World primates";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14536-14541(1997).  
 DR EMBL: U59637; AAB97480.1; -.  
 DR HSSP: P30685; IAY9.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 356 AA; 40093 MW; 728A4997 CRC32;  
 Query Match 100.0%; Score 51; DB 7; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 0.023; Mismatches 0; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RVDLRLTLCY 10  
 Db 91 RVDLRLTLCY 100  
 RESULT 7  
 Q31254 PRELIMINARY; PRT; 367 AA.  
 AC Q31254;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE MHC CLASS I RT1.AB HEAVY CHAIN PRECURSOR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CONA STIMULATED SPLENOCYTES;  
 RX MEDLINE: 96175158.  
 RA WANG M., STEPROWSKI S.M., TIAN L., LANGOWSKI J.L., YU J., KAHAN B.D.;  
 RT "Nucleotide sequences of three distinct clones coding for rat heavy  
 chain class I major histocompatibility antigens";  
 RL Immunogenetics 43:318-320(1996).  
 DR EMBL: U38970; AAC52550.1; -.  
 DR HSSP: P01900; 1B11.  
 DR PFAM: PF00047; ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW Signal; MHC.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 367 MHC CLASS I RT1.AB HEAVY CHAIN.  
 SQ SEQUENCE 367 AA; 41471 MW; 35A9D3A8 CRC32;

Query Match 100.0%; Score 51; DB 7; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 0.024; Mismatches 0; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RVDLRLTLCY 10  
 Db 96 RVDLRLTLCY 105  
 RESULT 8  
 Q19243 PRELIMINARY; PRT; 364 AA.  
 ID Q19243  
 AC Q19243  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE MHC CLASS I ANTIGEN.  
 GN PC14.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SULLIVAN J.A., OTTINGER H.F., SACHS D.H., EDGE A.S.B.;  
 RL J. Immunol. 159:0-0(1997).  
 DR EMBL: AF014004; AAB69339.1; -.  
 DR HSSP: P30685; IAY9.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC.  
 SQ SEQUENCE 364 AA; 40791 MW; BFC101BB CRC32;

Query Match 94.1%; Score 48; DB 7; Length 364;  
 Best Local Similarity 90.0%; Pred. No. 0.094; Mismatches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RVDLRLTLCY 10  
 Db 99 RVDLRLTLCY 108  
 RESULT 9  
 Q30221 PRELIMINARY; PRT; 356 AA.  
 AC Q30221;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE MHC CLASS I ATBE-B\*01 (FRAGMENT).  
 GN ATBE-B.  
 OS Ateles belzebuth (Long-haired spider monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98070787.  
 RA CADAVID L.F., SHUFFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.,  
 RA WATKINS D.I.;  
 RT "Evolutionary instability of the major histocompatibility complex  
 class I loci in New World primates";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14536-14541(1997).  
 DR EMBL: U59648; AAB97491.1; -.  
 DR HSSP: P01891; 2HLA.  
 DR PFAM: PF00047; ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW NON\_TER 1 1  
 SQ SEQUENCE 356 AA; 39667 MW; 7C371D68 CRC32;

Query Match 92.2%; Score 47; DB 7; Length 356;  
 Best Local Similarity 90.0%; Pred. No. 0.15; Mismatches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RVDLRLTLCY 10  
 Db 91 RVDLRLTLCY 100  
 RESULT 10  
 Q19744 PRELIMINARY; PRT; 90 AA.  
 ID Q19744  
 AC Q19744  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE MHC HLA-E ALPHA-1 (FRAGMENT).  
 GN EAL.2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PERIPHERAL BLOOD;  
 RA MEDLINE; 91033943.  
 RX OHYA K., KONDO K., MIZUNO S.;  
 RT "Polymorphism in the human class I MHC locus HLA-E in Japanese.";  
 RL Immunogenetics 32:205-209(1990).  
 DR EMBL; M32506; AAA63223.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 90 AA; 10505 MW; C19D8182 CRC32;

Query Match 90.28; Score 46; DB 7; Length 90;  
 Best Local Similarity 90.08; Pred. No. 0.054; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 1;

QY 1 RVDLRLTGRY 10  
 II:IIIIIIII  
 DB 75 RVDLRLTGRY 84

RESULT 11  
 O62875 PRELIMINARY; PRT; 144 AA.  
 AC O62875;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE MHC CLASS I ANTIGEN (FRAGMENT).  
 GN PIP1-E.  
 OS Pithecia pithecia (White-faced saki).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Platyrrhini; Cebidae; Pitheciinae; Pithecia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX KNAPP L.A., CADAVID L.F., WATKINS D.I.;  
 RT "The MHC-E locus is the most well conserved of all known primate class I histocompatibility genes";  
 RL J. Immunol. 160:189-196(1998).  
 DR EMBL; AF004922; AAC39680.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 144 AA; 16535 MW; DC03E510 CRC32;

Query Match 90.28; Score 46; DB 7; Length 144;  
 Best Local Similarity 90.08; Pred. No. 0.088; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 1;

QY 1 RVDLRLTGRY 10  
 II:IIIIIIII  
 DB 9 RVDLRLTGRY 18

RESULT 12  
 O46720 PRELIMINARY; PRT; 166 AA.  
 AC O46720;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE MHC CLASS I HEAVY CHAIN (FRAGMENT).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 DE MHC CLASS I HEAVY CHAIN (FRAGMENT).  
 RN [1]

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-302D;  
 RA URAKAWA T., KODAMA M., MORITA M., IKEDA H.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB008586; BAA24716.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 166 AA; 19338 MW; F0563A5A CRC32;

Query Match 90.28; Score 46; DB 7; Length 166;  
 Best Local Similarity 90.08; Pred. No. 0.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 0;

QY 1 RVDLRLTGRY 10  
 II:IIIIIIII  
 DB 70 RVDLRLTGRY 79

RESULT 13  
 O46722 PRELIMINARY; PRT; 166 AA.  
 AC O46722;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE MHC CLASS I HEAVY CHAIN (FRAGMENT).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-302D;  
 RA URAKAWA T., KODAMA M., MORITA M., IKEDA H.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB008589; BAA24719.1; -  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 166 AA; 19337 MW; 9321E1A4 CRC32;

Query Match 90.28; Score 46; DB 7; Length 166;  
 Best Local Similarity 90.08; Pred. No. 0.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 0;

QY 1 RVDLRLTGRY 10  
 II:IIIIIIII  
 DB 70 RVDLRLTGRY 79

RESULT 14  
 O46734 PRELIMINARY; PRT; 166 AA.  
 AC O46734;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE MHC CLASS I HEAVY CHAIN (FRAGMENT).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 OC Bovinae; Bos.  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-303F;  
RA URAKAWA T., KODAMA M., MORITA M., IKEDA H.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008605; BAA24734.1; -  
DR PFAM; PF00129; MHC\_I; 1.  
KW MHC.

FT NON\_TER 1

FT NON\_TER 166

SQ SEQUENCE 166 AA; 19338 MW; F0563A5A CRC32;

Query Match

Best Local Similarity 90.2%; Score 46; DB 7; Length 166;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10

DB 70 RVDLNLRLGY 79

RESULT 15

O46743

ID O46743 PRELIMINARY; PRT; 166 AA.

AC O46743;

DT 01-JUN-1998 (TREMELrel. 06, Created)

DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)

DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)

DE MHC CLASS I HEAVY CHAIN (FRAGMENT).

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovinae; Bos.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-303F;

RA URAKAWA T., KODAMA M., MORITA M., IKEDA H.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB008616; BAA24745.1; -

DR PFAM; PF00129; MHC\_I; 1.

KW MHC.

FT NON\_TER 1

FT NON\_TER 166

SQ SEQUENCE 166 AA; 19236 MW; 0B5B8001 CRC32;

Query Match

Best Local Similarity 90.2%; Score 46; DB 7; Length 166;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RVDLRLRGY 10

DB 70 RVDLNLRLGY 79

Search completed: February 8, 2000, 19:16:20

Job time: 21507 sec

**THIS PAGE BLANK (USPTO)**

Sequence	Strid	Orig	zScore	EScore	Len	Documentation
gb_ro:RNR1A2P	+	51.00	180.81	0.0476	1035	A13579 Rattus norvegicus
gb_ro:RNO249705	+	51.00	180.79	0.0478	1038	A249705 Rattus norvegicus
gb_ro:RNR1A1G	+	51.00	180.79	0.0478	1038	X30373 R.norvegicus mRNA for R
gb_ro:RNR1A20G	+	51.00	180.79	0.0478	1038	X30372 R.norvegicus mRNA for R
gb_ro:RNY1A014	+	51.00	180.76	0.0479	1041	A14014 Rattus norvegicus mRNA
gb_ro:RNPPEPPES	+	51.00	180.72	0.0482	1047	A08532 R.norvegicus mRNA for P
gb_ro:RNO249704	+	51.00	180.65	0.0486	1056	A249704 Rattus norvegicus
gb_pr3:ICJ059637	+	51.00	180.54	0.0493	1071	U59637 Gallitrix jacchus MHC
gb_ro:RNO38970	+	51.00	177.40	0.0738	1588	U38970 Rattus norvegicus MHC C
gb_ro:RATHWCKT	+	53.00	177.39	0.0739	1590	A1018 Rattus norvegicus MHC C
gb_om:AF014004	+	48.00	168.30	0.2369	1095	AF014004 Sus scrofa MHC class
gb_pr3:ABU59648	+	47.00	164.46	0.3879	1071	U59648 Ateles belzebuth MHC cl
gb_pr1:HSHLAE012	+	46.00	171.90	0.1493	254	X87678 H.sapiens HLA-E*01 varia
gb_pr1:HSHLAE1E2	+	46.00	171.90	0.1493	254	X87680 H.sapiens HLA-E*01 varia
gb_om:AB008253	+	46.00	171.41	0.1589	270	A002533 Homo sapiens HLA-E gen
gb_pr1:HSUMH2E11	+	46.00	171.41	0.1589	270	M32505 Human (Mongoloid) lymph
gb_pr1:HSUMH2E12	+	46.00	171.41	0.1589	270	M32506 Human (Mongoloid) MHC H
gb_om:BTMHC2	+	46.00	169.37	0.2066	349	X97645 B.taurus MHC class I gen
gb_pr3:AF004922	+	45.00	167.65	0.2575	433	AF004922 Pithecia pithecia MHC
gb_om:AB008586	+	45.00	166.54	0.2970	498	AB008586 Bos taurus mRNA for M
gb_om:AB008589	+	46.00	166.54	0.2970	498	AB008589 Bos taurus mRNA for M
gb_om:AB008616	+	46.00	166.54	0.2970	498	AB008616 Bos taurus mRNA for M
gb_om:AB008594	+	46.00	166.49	0.2989	501	AB008594 Bos taurus mRNA for M
gb_om:AB008622	+	46.00	166.49	0.2989	501	AB008622 Bos taurus mRNA for M
gb_om:AB008605	+	45.00	163.88	0.4176	695	AB008605 Bos taurus gene for M
gb_om:AB008609	+	45.00	163.86	0.4188	697	AB008609 Bos taurus gene for M
gb_om:AB008620	+	46.00	163.86	0.4188	697	AB008620 Bos taurus gene for M
gb_om:AB008621	+	46.00	163.86	0.4188	697	AB008621 Bos taurus gene for M
gb_om:AB008619	+	46.00	163.84	0.4200	699	AB008619 Bos taurus gene for M
gb_om:AB008625	+	46.00	163.84	0.4200	699	AB008625 Bos taurus gene for M
gb_om:AB008600	+	46.00	163.81	0.4212	701	AB008600 Bos taurus gene for M
gb_pr2:SGU52118	+	46.00	162.41	0.5043	836	U52118 Saguinus Geoffroyi MHC c
gb_pr2:SFU52117	+	46.00	162.37	0.5068	840	U52117 Saguinus fuscicollis MHC
gb_pr2:SMU52120	+	46.00	162.37	0.5068	840	U52120 Saguinus mystax MHC clas
gb_pr1:HSHLAEPX23	+	46.00	162.00	0.5314	880	X64880 H.sapiens gene for HLA-E
gb_ro:RNR1A1U	+	46.00	160.71	0.6273	1035	X82106 R.norvegicus mRNA for t
gb_ro:RNO249702	+	46.00	160.69	0.6291	1038	A249702 Rattus norvegicus
gb_ro:RNR1A1CG	+	46.00	160.69	0.6291	1038	X30370 R.norvegicus mRNA for R
gb_pr3:LRU59642	+	46.00	160.50	0.6440	1082	U59642 Leontopithecus rosalia
gb_pr3:LRU59643	+	46.00	160.50	0.6440	1082	U59643 Leontopithecus rosalia
gb_pr1:GGGOGA03	+	46.00	160.48	0.6458	1085	X54375 G.gorilla MHC class I A
gb_ro:RNO249700	+	46.00	160.45	0.6483	1089	A249700 Rattus norvegicus

```

|||||
223 CGAGTGGACCTGAGGACCTCGCGGCTAC 252
seq_name: gb_ro:RNO249705

seq_documentation_block:
LOCUS      RNO249705      1038 bp      mRNA      29-NOV-1999
DEFINITION Rattus norvegicus partial mRNA for MHC class Ia A2b antigen
            (Rtl-A2b gene).
ACCESSION  AJ249705
VERSION    AJ249705.1 GI:5912595
KEYWORDS   major histocompatibility complex; MHC class IA antigen; Rtl-A2b
            gene.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1038)
AUTHORS     Le Rolle, A., Joly, E. and Butcher, G.
TITLE       Characterisation of the rat MHC class I molecules
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1038)
AUTHORS     Le Rolle, A.F.
TITLE       Direct Submission
JOURNAL     Submitted (16-SEP-1999) Le Rolle A.F., Cellular Immunology Program,
            The Babraham Institute, CB2 4AF, Cambridge, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..1038
                     /organism="Rattus norvegicus"
                     /strain="BUF"
                     /db_xref="taxon:10116"
                     /chromosome="20"
                     /haplotype="Rtlb"
                     /clone="Eb+8"
     gene            1..1038
                     /gene="Rtl-A2b"
                     <1..1038
     CDS             /gene="Rtl-A2b"
                     /standard_name="MHC class Ia antigen"
                     /function="antigen presentation"
                     /codon_start=1
                     /product="A2b"
                     /protein_id="CAB56216.1"
                     /db_xref="GI:5912596"
                     /translation="GSHSLRYFETAVSRPGLCEPRYMEVGYVDDTQFVRYDSDAENPR
                     MEPRVQWIEREGREYWEQTONAKGEQVNRVDLRLRGYVNOSEGSHTIORMHGCD
                     VGTDSLRLRGYEHAYDGRDYIALNEDLKTWTAADFAAQITRNKWDQAGAAERSRAYL
                     EGTCEWLRILEHGRETLRLSDPPPEAHVTLHPRPGGVTLRCWALGFYPADITLTWQ
                     LNGEDLTQDMELVETRPAGDGTQKAWVVVPLGKEQNTCRVEHGLPEPLSQRWEP
                     SPSTNSNMETTVIYVVLGALIGTLAIGIVAVVRRRRRTGKGNYAPAGRDSSQ
                     SSDVSLPDCKA"
BASE COUNT      230 a      280 c      354 g      174 t
ORIGIN

alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-653-294-37 x RNO249705 ..
Align seg 1/1 to: RNO249705 from: 1 to: 1038
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||
223 CGAGTGGACCTGAGGACCTCGCGGCTAC 252
seq_name: gb_ro:RNR1A10G

seq_documentation_block:
LOCUS      RNR1A10G      1038 bp      mRNA      13-FEB-1998

```

```

DEFINITION R.norvegicus mRNA for Rtl.A1(O) alpha chain.
ACCESSION  X90373
VERSION    X90373.1 GI:1871632
KEYWORDS   alpha chain; class I antigen; major histocompatibility complex.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1038)
AUTHORS     Joly, E., Gonzalez, A.L. and Butcher, G.W.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1038)
AUTHORS     Joly, E.
TITLE       Direct Submission
JOURNAL     Submitted (27-JUL-1995) E. Joly, The Babraham Institute, Babraham,
            Cambridge, CB2 4AT, UK
REFERENCE   3 (bases 1 to 1038)
AUTHORS     Joly, E., Le Rolle, A.F., Gonzalez, A.L., Mehling, B., Stevens, J.,
            Coadwell, W.J., Huenig, T., Howard, J.C. and Butcher, G.W.
TITLE       Co-evolution of rat tap transporters and MHC class I Rtl-A
            molecules
JOURNAL     Curr. Biol. In press
COMMENT     On Mar 10, 1997 this sequence version replaced gi:940822.
FEATURES             Location/Qualifiers
     source          1..1038
                     /organism="Rattus norvegicus"
                     /strain="Rtl(O)"
                     /isolate="88(1) 29.6.92"
                     /db_xref="taxon:10116"
                     /chromosome="20"
                     /germline
                     /sex="Female"
                     /tissue.type="spleen"
                     /cell_type="splenocyte"
     gene            1..1038
                     /gene="Rtl.A1(O)"
                     <1..1038
     CDS             /gene="Rtl.A1(O)"
                     /codon_start=1
                     /product="Mature alpha chain of major histocompatibility
                     complex class I antigen"
                     /protein_id="CAA62023.1"
                     /db_xref="GI:1871633"
                     /db_xref="SPTREMBL:P79599"
                     /translation="GSHSLRYFETAVSRPGLCEPRYMEVGYVDDTQFVRYDSDAENPR
                     MEPRVQWIEREGREYWEQTONAKGEQVNRVDLRLRGYVNOSEGSHTIORMHGCD
                     VGTDSLRLRGYEHAYDGRDYIALNEDLKTWTAADFAAQITRNKWDQAGAAERSRAYL
                     EGTCEWLRILEHGRETLRLSDPPPEAHVTLHPRPGGVTLRCWALGFYPADITLTWQ
                     LNGEDLTQDMELVETRPAGDGTQKAWVVVPLGKEQNTCRVEHGLPEPLSQRWEP
                     SPSTNSNMETTVIYVVLGALIGTLAIGIVAVVRRRRRTGKGNYAPAGRDSSQ
                     SSDVSLPDCKA"
BASE COUNT      224 a      277 c      352 g      185 t
ORIGIN

alignment_scores:
  Quality: 51.00      Length: 10
  Ratio: 5.100      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-653-294-37 x RNR1A10G ..
Align seg 1/1 to: RNR1A10G from: 1 to: 1038
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||
223 CGAGTGGACCTGAGGACCTCGCGGCTAC 252
seq_name: gb_ro:RNR1A20G

seq_documentation_block:
LOCUS      RNR1A20G      1038 bp      mRNA      12-FEB-1998

```



```

DEFINITION R.norvegicus mRNA for RT1.A2(O) alpha chain.
ACCESSION X90372
VERSION X90372.1 GI:1871636
KEYWORDS alpha chain; class I antigen; major histocompatibility complex.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1038)
AUTHORS Joly,E., Gonzalez,A.L. and Butcher,G.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1035)
AUTHORS Joly,E.
JOURNAL Direct Submission
TITLE Submitted (27-JUL-1995) E. Joly, The Babraham Institute, Babraham,
Cambridge, CB2 4AT, UK
REFERENCE 3 (bases 1 to 1038)
AUTHORS Joly,E., Le Rolle,A.F., Gonzalez,A.L., Mehling,B., Stevens,J.,
Coadwell,W.J., Huenig,T., Howard,J.C. and Butcher,G.W.
TITLE Co-evolution of rat tap transporters and MHC class I RT1-A
molecules
JOURNAL Curr. Biol. In press
COMMENT On Mar 10, 1997 this sequence version replaced gi:940828.
FEATURES
    source
        1..1038
            /organism="Rattus norvegicus"
            /strain="RT1(O)"
            /isolate="88(1) 29.6.92"
            /db_xref="taxon:10116"
            /chromosome="20"
            /germline
            /sex="Female"
            /tissue_type="spleen"
            /cell_type="splenocyte"
            /gene="RT1.A2(O)"
            1..1038
            /gene="RT1.A2(O)"
            /codon_start=1
            /product="Mature alpha chain of major histocompatibility
            complex class I antigen"
            /protein_id="CAA62022.1"
            /db_xref="GI:1871637"
            /translation="GSHSLRYFYIAVSRPGLGEPRIYMEVGVYDDTFQVRYDSDAENPR
            YEPARWMEGPEYWEETQAKGEQIYVDLRLRGYINQSEGGSHITQEMYGCD
            VSGDGLRGYQDAYDGRDYIALNEDLKTAAADFAAQITRNKWRARYAERLRLAYL
            EGTGVESLRLYLELKEKTLKSDPPEAHVTLHPRPGDVTLCRCWALGFYPADITLTWQ
            LNGEDLTQDMELVETRPAGDGTFOKWSVYVPLGKEQNTCRVEHGLPEPLSORWEP
            SPSTNSNMETTIVYVLLGAIIGTILAIIGIVAVVRRRRNTGKGGNYAPAPGRDSSQ
            SSDVSLPCKA"
            226 a 278 c 352 g 182 t
BASE COUNT
ORIGIN

alignment_scores:
    Quality: 51.00 Length: 10
    Ratio: 5.100 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-653-294-37 x RNRT1A20G ..
Align seg 1/1 to: RNRT1A20G from: 1 to: 1038
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||
223 CGAGTGACCTGAGGACCTCGCGCGCTAC 252
seq_name: gb_ro:RN14014
seq_documentation_block:
LOCUS RN14014 1041 bp mRNA ROD 12-FEB-1998

```

```

DEFINITION Rattus norvegicus mRNA for MHC class I alpha chain, partial.
ACCESSION Y14014
VERSION Y14014.1 GI:2225994
KEYWORDS MHC class I alpha chain.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Mehling,B.
JOURNAL Direct Submission
TITLE Submitted (17-JUN-1997) Mehling B., Universitaet Wuerzburg,
Institut fuer Virologie und Immunbiologie, Versbacherstr. 7 97078
Wuerzburg, GERMANY
REFERENCE 2 (bases 1 to 1041)
AUTHORS Joly,E., LeRolle,A., Gonzalez,A.L., Mehling,B., Stevens,J.,
Coadwell,W.J., Huenig,T., Howard,J.C. and Butcher,G.W.
TITLE Co-evolution of rat TAP transporters and MHC class I RT1-A
molecules
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1041)
AUTHORS Joly,E., Le Rolle,A.F., Gonzalez,A.L., Mehling,B., Stevens,J.,
Coadwell,W.J., Huenig,T., Howard,J.C. and Butcher,G.W.
TITLE Co-evolution of rat tap transporters and MHC class I RT1-A
molecules
JOURNAL Curr. Biol. In press
FEATURES
    source
        1..1041
            /organism="Rattus norvegicus"
            /strain="LEW.lf (f haplotype)"
            /db_xref="taxon:10116"
            /chromosome="20"
            /sex="Female"
            /tissue_type="spleen"
            /cell_type="splenocyte"
            /gene="RT1.A2(O)"
            1..1038
            /gene="RT1.A2(O)"
            /codon_start=1
            /product="MHC class I alpha chain"
            /protein_id="CAA74333.1"
            /db_xref="GI:2225995"
            /translation="GSHSLRYFYIAVSRPGLGEPRIAGVVDTEFVRFDSDAENPR
            MEPRARWMEGPEYWEETQAKGEQIYVDLRLRGYINQSEGGSHITQEMYGCD
            VSGDGLRGYQDAYDGRDYIALNEDLKTAAADFAAQITRNKWRARYAERLRLAYL
            EGTGVESLRLYLELKEKTLKSDPPEAHVTLHPRPGDVTLCRCWALGFYPADITLTWQ
            LNGEDLTQDMELVETRPAGDGTFOKWSVYVPLGKEQNTCRVEHGLPKPLSORWEP
            SPSTDSNMETTIVYVLLGAVAMIGAVIIGAVVAVRRKRNTGKGEYAPAPGRDSS
            QSSDVSLPCKA"
            271..546
            /note="alpha 2-domain"
            /number=3
            547..822
            /note="alpha 3-domain"
            /number=4
            823..957
            /note="transmembrane"
            /number=5
            958..987
            /note="cytoplasmic"
            /number=6
            988..1041
            /note="cytoplasmic"
            /number=7
BASE COUNT 226 a 281 c 349 g 185 t
ORIGIN

```

```

alignment_scores:
    Quality: 51.00 Length: 10
    Ratio: 5.100 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-653-294-37 x RNY14014 ..

```

Align seg 1/1 to: RNY14014 from: 1 to: 1041

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
|||||  
223 CGAGTGGACCTGAGGACCTCGCGGCTAC 252

seq\_name: gb\_ro:RNPEPPRES

seq\_documentation\_block:  
LOCUS RNPEPPRES 1047 bp mRNA ROD 12-FEB-1998  
DEFINITION R.norvegicus mRNA for protein involved in peptide presentation.  
ACCESSION Y08532  
VERSION Y08532.1 GI:1871630  
KEYWORDS peptide presentation protein.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 1047)  
AUTHORS Le Rolle,A.F., Butcher,G.W. and Joly,E.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1047)  
AUTHORS Joly,E.  
TITLE Direct Submission  
JOURNAL Submitted (01-OCR-1996) E. Joly, The Babraham Institute,  
Immunology, Babraham Hall, Babraham, Cambridge CB2 4AT, UK  
REFERENCE 3 (bases 1 to 1047)  
AUTHORS Joly,E., Gonzalez,A.L. and Butcher,G.W.  
JOURNAL Unpublished  
REFERENCE 4 (bases 1 to 1047)  
AUTHORS Joly,E., Le Rolle,A.F., Gonzalez,A.L., Mehling,B., Stevens,J.,  
Coadwell,W.J., Huenig,F., Howard,J.C. and Butcher,G.W.  
TITLE Co-evolution of rat tap transporters and MHC class I RT1-A  
molecules  
JOURNAL Curr. Biol. In press  
COMMENT On Mar 10, 1997 this sequence version replaced gi:1619331.  
FEATURES  
Location/Qualifiers  
1..1047  
/organism="Rattus norvegicus"  
/strain="NEDH"  
/db\_xref="taxon:10116"  
/chromosome="20"  
/germline  
/haplotype="RT1g"  
/tissue\_type="splenocytes"  
/note="identical sequence obtained from PC12 cells"  
1..1038  
/gene="RT1.Ag"  
/db\_xref="SPRMBL:P70527"  
/translation="GSHSLRYFLTVASRPGLPGEPIFYVDDTEFVYDSDAENPR  
MEPRVQWIEREGPEYWEQTRAKGHEQVNRVDLTLRGYINQSEGSHTIQMFQCD  
VGTDSLLRGYRDADYGRDIYALNEDLKTWTAADFAAQITRSKWERYAERLRYAL  
EGECVGLRRLYLEHGKELLRSDDPPKKAHVTLHPRPGDVTLCRWALGFYPADITLWQ  
LNGLDITQDMELVETRPAGDTQKWAASVVVPLGKEONYTCRVEHGLPEPLSQRPWP  
PPSDSNMETIYIVVLGAVAIIAAVIIGAVVAVVRRRRRTGKGDTYTPAGTDSQ  
SSDVSPLDCKA"

BASE COUNT 230 a 275 c 350 g 192 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-653-294-37 x RNY14014 from: 1 to: 1041

US-08-653-294-37 x RNPEPPRES

Align seg 1/1 to: RNPEPPRES from: 1 to: 1047

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
|||||  
223 CGAGTGGACCTGAGGACCTCGCGGCTAC 252

seq\_name: gb\_ro:RNO249704

seq\_documentation\_block:  
LOCUS RNO249704 1056 bp mRNA ROD 29-NOV-1999  
DEFINITION Rattus norvegicus partial mRNA for MHC class Ia Alb antigen  
(RT1-Alb gene).  
ACCESSION AJ249704  
VERSION AJ249704.1 GI:5912593  
KEYWORDS major histocompatibility complex; MHC class Ia antigen; RT1-Alb  
gene.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 1056)  
AUTHORS Le Rolle,A., Joly,E. and Butcher,G.  
TITLE Characterisation of the rat MHC class I molecules  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1056)  
AUTHORS Le Rolle,A.F.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-1999) le Rolle A.F., Cellular Immunology Program,  
The Babraham Institute, CB2 4AF, Cambridge, UNITED KINGDOM  
FEATURES  
Location/Qualifiers  
1..1056  
/organism="Rattus norvegicus"  
/strain="BUF"  
/db\_xref="taxon:10116"  
/chromosome="20"  
/haplotype="RT1b"  
/clone="Ba8"  
1..1041  
/gene="RT1-Alb"  
/db\_xref="GI:5912594"  
/translation="GSHSLRYFLTVASRPGLPGEPIFYVDDTEFVYDSDAENPR  
YEPRAWMEREGPEYWEQTRAKENEQVNRVDLTLRGYINQSEGSHTIQMFQCD  
VSDGSLRGYEQHAYDGRDIYALNEDLKTWTAADFAAQITRSKWERYAERLRYAL  
EGTCVEMLRRLYLEHGKELLRSDDPPKKAHVTLHPRPGDVTLCRWALGFYPADITLWQ  
LNGLDITQDMELVETRPAGDTQKWAASVVVPLGKEONYTCRVEHGLPEPLSQRPWP  
SLSDSNMETIYIVVLGAVAIIAAVIIGAVVAVVRRRRRTGKGDTYTPAGTDSQ  
QSSDVSPLDCKA"

BASE COUNT 226 a 287 c 354 g 189 t  
ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 10  
Ratio: 5.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-653-294-37 x RNO249704

Align seg 1/1 to: RNO249704 from: 1 to: 1056

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
|||||  
223 CGAGTGGACCTGAGGACCTCGCGGCTAC 252

```

seq_name: gb_pr3:CUJ59637
seq_documentation_block: 1071 bp mRNA PRI 22-JAN-1998
LOCUS CUJ59637
DEFINITION Callithrix jacchus MHC class I (Caja-G*01) mRNA, partial cds.
ACCESSION U59637
VERSION U59637.1 GI:1389920
KEYWORDS
SOURCE
ORGANISM Callithrix jacchus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
REFERENCE
1 (bases 1 to 1071)
AUTHORS Cadavid,L.F., Shuffelebotham,C., Ruiz,F.J., Yeager,M., Hughes,A.L.
and Watkins,D.I.
TITLE Evolutionary instability of the major histocompatibility complex
class I loci in new world primates
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (26), 14536-14541 (1997)
MEDLINE 98070787
REFERENCE
2 (bases 1 to 1071)
AUTHORS Cadavid,L.F., Shuffelebotham,C., Ruiz,F.J., Yeager,M., Hughes,A.L.
and Watkins,D.I.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-1996) Wisconsin Regional Primate Research Center,
University of Wisconsin-Madison, 1220 Capitol Court, Madison, WI
53715, USA
FEATURES
source
1..1071
/organism="Callithrix jacchus"
/db_xref="taxon:9483"
gene
1..1071
/organism="Callithrix jacchus"
CDS
<1..1071
/gene="MHC class I Caja-G"
/gene="MHC class I Caja-G"
/note="Allele: 01"
/codon_start=1
/product="MHC class I Caja-G*01"
/protein_id="AAB97480.1"
/db_xref="GI:1389921"
/translaton="LLLLLGVLTWAGSHSMRYFYSVSRPGRGPRFIIVGV
DDTFVRFDSANPRPEPRAPWVEQGEYWEQTRAKAHQOTDVRDLRLGYIN
QSDAGSTIQMSSCVVPGDRLRGYQDAYDQKDYIALNEDLRSWTAADMAQITQ
RKWEAANAEMRAYLGECELEWLRHYLENAKETLQRAEPPKTHVTHHPVSDHEATLR
CWALGFPAETTLTWQDGEDOTODMELVETPTDRTFOKWAAVVPSGEEHRYTCH
VQHEGLPEPLTRHEPFSQPTIPINGIVAALAILGAVGGVAVVAVMVRKXSKKGGG
YSQAARSESAGGSDVSUTACKV"
BASE COUNT 214 a 321 c 366 g 170 t
ORIGIN

alignment_scores:
Quality: 51.00 Length: 10
Ratio: 5.100 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-37 x CUJ59637
..
Align seg 1/1 to: CUJ59637 from: 1 to: 1071
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||
271 CGAGTGGACCTGCGGACCTCGCGGCTAC 300

seq_name: gb_ro:RNU38970
seq_documentation_block: 1588 bp mRNA ROD 16-MAY-1996
LOCUS RNU38970
DEFINITION Rattus norvegicus MHC class I RT1.Ab heavy chain precursor, mRNA,
complete cds.
ACCESSION U38970
VERSION U38970.1 GI:1263197
KEYWORDS
SOURCE
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 1590)
AUTHORS Wang,M., Stepkowski,S.M., Tian,L., Langowski,J.L., Yu,J. and
Kahan,B.D.
TITLE Nucleotide sequences of three distinct clones coding for rat heavy
chain class I major histocompatibility antigens
JOURNAL Immunogenetics 43 (5), 318-320 (1996)
MEDLINE 96175158
REFERENCE
2 (bases 1 to 1588)
AUTHORS Wang,M., Tian,L., Langowski,J.L., Stepkowski,S.M. and Kahan,B.D.
Direct Submission
TITLE Submitted (20-OCT-1995) M. Wang, Division of Immunology & Organ
Transplantation, Department of Surgery, University of Texas Medical
School, 6431 Fannin Street MSB 6.255, Houston, TX 77030, USA
FEATURES
source
1..1588
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/haplotype="RT1.Ab"
/tissue_type="ConA stimulated splenocytes"
1..63
sig_peptide
1..1104
CDS
1..1104
/codon_start=1
/product="MHC class I RT1.Ab heavy chain precursor"
/protein_id="AAC52550.1"
/db_xref="GI:1263198"
/translaton="MAPRTLLLLAALAPTQTRAGSHSLRYFYTAVSRPLGEPRI
SGYVDDTEFVYDSDAENPRYEPRAEMEREGYWEQOTRIAKENEQVNRVDLRTL
RGYVNOSEGSSTIQEMFGCDVSGSLRGYEQHAYDGRDYIALNEDLKTWTAADFA
AQITRSKWARARYAERLRLAYLESTCVENLRYLEHSGKETLLRSDDPKAHVTPRPREG
DVTLCWALGFYADISLWQLNGEDLTQDMELVETRPAGDGFQKWSAVVPLKREQ
NYTCRVEHEGLPEPLSRQWEPSTLSNMETYYIVYVVLGAVIAAIVAAVTVVRR
RRRTGKGKGVYTPAPGRDSSQSSVSLPDKA"
BASE COUNT 346 a 426 c 484 g 332 t
ORIGIN

alignment_scores:
Quality: 51.00 Length: 10
Ratio: 5.100 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-653-294-37 x RNU38970
..
Align seg 1/1 to: RNU38970 from: 1 to: 1588
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||
286 CGAGTGGACCTGAGGACCTCGCGGCTAC 315

seq_name: gb_ro:RATMHCRT
seq_documentation_block: 1590 bp mRNA ROD 11-MAR-1997
LOCUS RATMHCRT
DEFINITION Rattus norvegicus MHC class I RT1.Aa alpha-chain precursor mRNA,
complete cds.
ACCESSION M31018
VERSION M31018.1 GI:1877415
KEYWORDS RT1.Aa alpha-chain; cell surface antigen; cell surface
glycoprotein; class I gene; integral membrane protein; major
histocompatibility complex.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 1590)

```

AUTHORS Rada,C., Lorenzi,R., Powis,S.J., van den Bogaerde,J., Parham,P. and Howard,J.C.  
 TITLE Concerted evolution of class I genes in the major histocompatibility complex of murine rodents  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (6), 2167-2171 (1990)  
 MEDLINE 90192768  
 REFERENCE 2 (bases 1 to 1590)  
 AUTHORS Howard,J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-DEC-1990) The Institute for Genetics, University of Cologne, Cologne, 50674, Germany  
 REFERENCE 3 (bases 1 to 1590)  
 AUTHORS Howard,J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-MAR-1997) The Institute for Genetics, University of Cologne, Cologne, 50674, Germany  
 REMARK Sequence update by submitter  
 COMMENT On Mar 11, 1997 this sequence version replaced gi:205421.  
 FEATURES  
 source  
 1..1590  
 /organism="Rattus norvegicus"  
 /strain="DA"  
 /db\_xref="taxon:10116"  
 /clone="3.3/1"  
 /sex="male"  
 /tissue\_type="spleen and lymph node"  
 /cell\_line="40 hour ConA blast cell line"  
 10..1125  
 /notes="5' end could be at nucleotide 19"  
 /codon\_start=1  
 /product="MHC R1.Aa alpha-chain precursor"  
 /protein\_id="AB49324.1"  
 /db\_xref="GI:1877416"  
 CDS  
 10..81  
 /translation="MEMAPRTLLLLAALAPTQTTRAGSHSLRYFYTAVSRLGPEP  
 RTIAGVVDTEFRFDSAPNPRMREGEPEYEQQTRIAKEWQIYRVDL  
 RTLRGYNQSEGSHFTQEMGCDVSGDGLLRGYRDAYDGRDVIADNEDLKTWAA  
 DPAQOITRNKRAERAYELRAVLEGTCCVEMLSVLELKEKTLASDPPEAHVTLPR  
 PGQDYLRCWALGFYPADTLTLQNLGEDLTQDMELVETRPAGDGTFOKWAAYVPLG  
 KEQYTCVRHEGLPPLSQRWPEPSTDSNMETIYVILGAVAMIGAVAILGAWVA  
 VVRRKRTGGGGYAPAPGRDSSQSDVSLPDCKA"  
 sig\_peptide  
 82..1122  
 /product="MHC R1.Aa alpha chain signal peptide"  
 mat\_peptide  
 158..11564  
 /products="MHC R1.Aa alpha chain"  
 polyA\_signal  
 336 a 423 c 493 g 338 t  
 BASE COUNT  
 ORIGIN  
 alignment\_scores  
 Quality: 51.00 Length: 10  
 Ratio: 5.100 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block  
 US-08-653-294-37 x RATWHCRT ..  
 Align seg 1/1 to: RATWHCRT from: 1 to: 1590  
 1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
 |||||  
 304 CGAGTGACCTGAGGACCTCGCGGGCTAC 333  
 seq\_name: gb\_om:AF014004  
 seq\_documentation\_block:  
 LOCUS AF014004 1095 bp mRNA MAM 03-SEP-1997  
 DEFINITION Sus scrofa MHC class I antigen (PC14) mRNA, complete cds.  
 ACCESSION AF014004  
 VERSION AF014004.1 GI:23252985  
 KEYWORDS pig.  
 SOURCE Sus scrofa  
 ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 REFERENCE  
 1 (bases 1 to 1095)  
 AUTHORS Sullivan,J.A., Oettinger,H.F., Sachs,D.H. and Edge,A.S.B.  
 TITLE Analysis of polymorphism in porcine MHC class I genes: Alterations  
 in signals recognized by human cytotoxic lymphocytes  
 J. Immunol. 159 (1997) In press  
 REFERENCE 2 (bases 1 to 1095)  
 AUTHORS Edge,A.S.B.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-1997) Molecular and Cellular Biology, Diacrin  
 Inc., Building 96, 13th Street, Charlestown, MA 02129, USA  
 FEATURES  
 source  
 1..1095  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 1..1095  
 /gene="PC14"  
 1..1095  
 /gene="PC14"  
 /note="SLA"  
 /codon\_start=1  
 /product="MHC class I antigen"  
 /protein\_id="AAB69339.1"  
 /db\_xref="GI:2352986"  
 /translation="MRVRGQAILLLSGALATGTRAGPHSLSYFYTAVSRLPDLGDS  
 RFIAGVYDDTFVRFDSAPNPRMREGEPEYEQQTRIAKEWQIYRVDL  
 KTLRGYNQSEGSHFTQEMGCDVSGDGLLRGYRDAYDGRDVIADNEDLKTWAA  
 DMAQISKRKEAANAQAQERSYLGRCVEMLSVLELKEKTLQRAEPPKTHVTRHPS  
 SDLGYTLRCWALGFYPKEISLTWOREGQSDMELVETRPAGDGTFOKWAALVVPVG  
 EEOSYTCVHQHEGLQEPILTLEWDPQPPVPIVGLVLVAGAMVAGVYVWRKKR  
 SGEKGGSTQAAGSDSAQSDVSLTKDRV"  
 BASE COUNT 211 a 340 c 368 g 176 t  
 ORIGIN  
 alignment\_scores  
 Quality: 48.00 Length: 10  
 Ratio: 4.800 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 90.000  
 alignment\_block  
 US-08-653-294-37 x AF014004 ..  
 Align seg 1/1 to: AF014004 from: 1 to: 1095  
 1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
 |||||  
 295 CGAGTGACCTGAAGACCTCGCGGGCTAC 324  
 seq\_name: gb\_pr3:ABU59648  
 seq\_documentation\_block:  
 LOCUS ABU59648 1071 bp mRNA PRI 22-JAN-1998  
 DEFINITION Ateles belzebuth MHC class I (Atbe-B01) mRNA, partial cds.  
 ACCESSION U59648  
 VERSION U59648.1 GI:1389908  
 KEYWORDS  
 SOURCE long-haired spider monkey.  
 ORGANISM Ateles belzebuth  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Platyrrhini; Cebidae; Ateles.  
 REFERENCE 1 (bases 1 to 1071)  
 AUTHORS Cadavid,L.F., Shuffiebotham,C., Ruiz,F.J., Yeager,M., Hughes,A.L.  
 and Watkins,D.I.  
 TITLE Evolutionary instability of the major histocompatibility complex  
 class I loci in new world primates  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (26), 14536-14541 (1997)  
 MEDLINE 98070787  
 REFERENCE 2 (bases 1 to 1071)  
 AUTHORS Cadavid,L.F., Shuffiebotham,C., Ruiz,F.J., Yeager,M., Hughes,A.L.  
 and Watkins,D.I.  
 TITLE Direct Submission

**JOURNAL** Submitted (31-MAY-1996) Wisconsin Regional Primate Research Center,  
University of Wisconsin-Madison, 1220 Capitol Court, Madison, WI  
53715, USA

FEATURES	source	Location/Qualifiers
		1. .1071
		/organism="Ateles belzebuth"
		/db_xref="taxon:9507"
gene		1. .1071
		/gene="Atbe-B"
CDS		<1..1071
		/gene="Atbe-B"
		/note="Allele: 01"
		/codon_start=1
		/product="MHC class I Atbe-B*01"
		/protein_id="AAB97491.1"
		/db_xref="GI:1389909"
		/translation="LLLLSGALALTQTRAGSHSMRYFVTSVRPGRGEPRTFVGVY DQTFVFRSDMAIPMEPRFLMWEEGPEYWEQTRRYKAAATDPRVLDLOTURGYN QSPAGSHLTQTMVYCDVDPGPRFLRGYRQDAYDGKDIALNEDLRSWTAADMAQNTK RKWEANVAEQLRAYLEGKQCSESLRYLENGKETLORADPPKTHVHPVSDHEALIR CNWAGLPETITLWQDGEDQDTLVEPTPAGDTFOKWAAYVVPSPGEGRYTCYI VQHEGPEITLWVPESSQLTIPVGIAGLVLAVVIGAAVTAVMWRRKSSGGKGS YSQAQSDSAQGSQDVSLTACKA"
BASE COUNT	216 a	319 c 173 t

```

alignment_scores:
  Quality: 47.00      Length: 10
  Ratio: 4.700       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 90.000

alignment_block:
  US-08-653-294-37 x ABUS9648 ..
  Align seg 1/1 to: ABUS9648 from: 1 to: 1071

      1 ArgValAspLeuArgThrLeuArgGlyTyr 10
      |||||
271 CGATTGGACCTGCAGACCTCGCGGCTAC 300

```

```

seq_name: gb_prl: HSHLAE012

seq_documentation_block:
LOCUS      HSHLAE012      254 bp      DNA
DEFINITION H.sapiens HLA-E*01 variant (HLA-E*01C230), exon 2.
ACCESSION  X87678
VERSION    X87678.1 GI:1109758
KEYWORDS   alpha chain; HLA-E*01; MHC class I.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 254)
AUTHORS   Blasczyk,R., Forstmann,G. and Salama,A.
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 254)
AUTHORS   Blasczyk,R.
TITLE     Direct Submission
JOURNAL   Submitted (31-MAY-1995) R. Blasczyk, Bloodbank, Dept of Internal
            Medicine, Div. of Hematology & Oncology, Virchow-Klinikum,
            Humboldt-Univ. Berlin, Augustenburger Platz 1, 13353 Berlin,
            Germany.
REFERENCE  3 (bases 1 to 254)
AUTHORS   Blasczyk,R.
TITLE     Direct Submission
JOURNAL   Submitted (07-DEC-1995) R. Blasczyk, Bloodbank, Dept of Internal
            Medicine, Div. of Hematology & Oncology, Virchow-Klinikum,
            Humboldt-Univ. Berlin, Augustenburger Platz 1, 13353 Berlin,
            Germany.
COMMENT    On Dec 8, 1995 this sequence version replaced gi:871291.
FEATURES   Location/Qualifiers
            1..254
                /organism="Homo sapiens"
                /isolate="940459"

```

```

/db_xref="taxon:9606"
/chromosome="6"
/map="p21.3"
/note="allele HLA-E*01C230"
17..254
/gene="HLA-E*01"
17..>254
/gene="HLA-E*01"
/number=2
/product="alpha chain MHC class I glycoprotein"
46 a 78 c 90 g 40 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 45.00      Length: 10
  Ratio: 4.500       Gaps: 0
  Percent similarity: 100.000  Percent identity: 90.000

alignment_block:
  US-08-653-294-37 x HSHLAE012

```

Align seg 1/1 to: HSHLAE012 from: 1 to: 254

```
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||:|||||
206 CGAGTGACCTCGGACGCTGGCGGCTAC 235
```

seq\_name: gb\_pr1:HSHLAE1E2

seq\_documentation\_block:

LOCUS	HSHLAE1E2	254 bp	DNA	PRI
DEFINITION	H.sapiens HLA-E*01 variant (HLA-E*01230), exon 2.			07-DEC-1995
ACCESSION	X87680			
VERSION	X87680.1	GI:1109760		
KEYWORDS	alpha chain; HLA-E*01; MHC class I.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			

RECEIVED	1995	Blaszyk, R., Forstmann, G. and Salama, A.
AUTHORS		Unpublished
JOURNAL		2 (bases 1 to 254)
REFERENCE		Blaszyk, R.
AUTHORS		Direct Submission
JOURNAL		Submitted (31-MAY-1995) R. Blaszyk, Bloodbank, Dept of Internal Medicine, Div. of Hematology & Oncology, Virchow-Klinikum, Humboldt-Univ. Berlin, Augustenburger Platz 1, 13353 Berlin, FRG
REFERENCE		3 (bases 1 to 254)
AUTHORS		Blaszyk, R.
JOURNAL		Direct Submission
TITLE		Submitted (07-DEC-1995) R. Blaszyk, Bloodbank, Dept of Internal Medicine, Div. of Hematology & Oncology, Virchow-Klinikum, Humboldt-Univ. Berlin, Augustenburger Platz 1, 13353 Berlin, FRG
COMMENT		On Dec 8, 1995 this sequence version replaced gi:871293.
FEATURES		Location/Qualifiers

```

i. .254
/organism="Homo sapiens"
/isolate="940637"
/db_xref="taxon:9606"
/chromosome="6"
/map="p21.3"
/note="allele HLA-E*01230"
17. .254
/gene="HLA-E*01"
17. .>254
/gene="HLA-E*01"
/number=3
/product="alpha chain MHC class I"
46 a 77 c 90 g 41 t
BASE COUNT
ORIGIN

```

alignment\_scores:  
Quality: 46.00 Length: 10  
Ratio: 4.600 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 90.000

## alignment\_block:

US-08-653-294-37 x HSHLAE1E2 ..  
Align seg 1/1 to: HSHLAE1E2 from: 1 to: 254

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
|||||:|||||:|||||:|||||:|||||  
206 CGAGTGAATCTGGGACGCTCGCGGCTAC 235

seq\_name: gb\_prl: HSAJ2533

seq\_documentation\_block:  
LOCUS HSAJ2533 270 bp DNA PRI 19-JAN-1999  
DEFINITION Homo sapiens HLA-E gene exon 2.  
ACCESSION AJ002533  
VERSION AJ002533.1 GI:3929738  
KEYWORDS HLA-E gene; human leukocyte antigen.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 270)  
Steffensen R.  
Direct Submission  
Submitted (31-OCT-1997) Steffensen R., Aalborg Sygehus, Department  
of Clinical Immunology and Blood Transfusion, 9000 Aalborg, DENMARK  
2 (bases 1 to 270)  
Steffensen R., Christiansen, O.B., Bennett, E.P. and Jersild, C.  
HLA-E polymorphism in patients with recurrent spontaneous abortion  
Tissue Antigens 52 (6), 569-572 (1998)  
99110147  
FEATURES  
Location/Qualifiers  
1..270  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..270  
/gene="HLA-E"  
/number=2  
1..270  
/gene="HLA-E"  
<1..>270  
/gene="HLA-E"  
/note="PCR product derived from human genomic DNA,  
confirmed by sequence analysis"  
/codon\_start=3  
/protein\_id="CAA05526.1"  
/db\_xref="GI:3929739"  
/translation="SHSLKYFHTSVSRPGRGEPFRFISGVYDDTQTFVRFNDNDAA SPRM  
VPRAPMEOEGSEYWDRETRSDAQIFRVNLTIRLGYNQSEA"

BASE COUNT  
ORIGIN

49 a 85 c 92 g 44 t

alignment\_scores:  
Quality: 46.00 Length: 10  
Ratio: 4.600 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 90.000

## alignment\_block:

US-08-653-294-37 x HSAJ2533 ..

Align seg 1/1 to: HSAJ2533 from: 1 to: 270

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
|||||:|||||:|||||:|||||:|||||  
222 CGAGTGAACCTGGGACGCTCGCGGCTAC 251

OM of: US-08-653-294-37 to: N\_Geneseq\_36:\* out\_format : pfs  
Date: Feb 8, 2000 7:32 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet.p2n.model -DEV=xlip  
-Q/cgn1\_1/USPTO\_spool/US08653294/runat\_04022000\_160701\_15807/app\_query.fasta.2  
-DB=N\_Geneseq\_36 -QFMT=fastap -SUFFIX=ring -GAPOB=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOB=4.500 -GAPEXT=0.050 -XGAPOB=10.000 -XGAPEXT=0.500  
-GAPEXT=6.000 -GAPEXT=7.000 -XGAPOB=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=ptt  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

## Search information block:

Query: US-08-653-294-37  
Query length: 10  
Database: N\_Geneseq\_36:\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 873.190000

## score\_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
N_Geneseq_36:V30457	+	44.00	143.87	2,06	Chimeric HLA-A2.1/beta-2 micro
N_Geneseq_36:Q12116	+	40.00	130.37	11,63	HLA-C exon Cb-1. HLA-C gene, D
N_Geneseq_36:Q21117	+	40.00	130.37	11,63	HLA-C exon Cb-2. HLA-C gene, D
N_Geneseq_36:V21187	+	38.00	89.07	2,3e+03	Amcyolaptosis mediterranei ri
N_Geneseq_36:V52205	-	37.00	98.60	684.11	Streptococcus pneumoniae gen
N_Geneseq_36:V52267	-	37.00	97.96	742.49	Streptococcus pneumoniae gen
N_Geneseq_36:T06769	-	37.00	90.86	1,8e+03	Sorangium cellulosum soraphen
N_Geneseq_36:T89956	-	37.00	90.75	1,9e+03	Sorangium cellulosum soraphen
N_Geneseq_36:V05287	-	37.00	86.10	3,4e+03	The soraphen biosynthesis gen
N_Geneseq_36:X32031	-	36.00	102.13	434.94	Swine HEV ORF 1 DNA. New isola
N_Geneseq_36:X32032	-	36.00	99.16	636.34	Swine HEV ORF 1 genomic DNA. N
N_Geneseq_36:X20248_06	-	36.00	75.42	1,3e+04	Continuation (7 of 10) of
N_Geneseq_36:X20248_07	-	36.00	75.42	1,3e+04	Continuation (8 of 10) of
N_Geneseq_36:T17935	-	35.00	121.56	35.98	Taenia ovis antigenic polypepti
N_Geneseq_36:Q27239	-	35.00	117.11	63.67	Taenia ovis antigenic polypepti
N_Geneseq_36:T17934	-	35.00	117.11	63.67	Taenia ovis antigenic polypepti
N_Geneseq_36:Q10571	-	35.00	112.39	116.71	Lipase modulating factor gene,
N_Geneseq_36:Q44282	-	35.00	106.12	260.62	Lipase/lipase modulator fusion
N_Geneseq_36:Q5131	+	35.00	97.76	762.12	Sequence encoding enzymes whic
N_Geneseq_36:Q5132	+	35.00	97.75	762.43	Sequence encoding enzyme which
N_Geneseq_36:T38971	+	35.00	97.75	762.43	Rhodococcus rhodochrous desulf
N_Geneseq_36:X15615	+	35.00	97.75	762.43	DNA encoding an enzyme capable
N_Geneseq_36:X15616	+	35.00	97.75	762.43	DNA encoding an enzyme capable
N_Geneseq_36:Q04525	+	35.00	69.95	2,6e+04	Total base sequence of rice
N_Geneseq_36:T90899	+	34.00	112.53	114.54	Von Hippel-Lindau disease gene
N_Geneseq_36:Q12115	+	34.00	108.21	199.43	HLA-B35 exon. HLA-B35 gene - u
N_Geneseq_36:Q78896	-	34.00	103.75	333.15	VHL disease gene 97. New phage
N_Geneseq_36:T90895	-	34.00	103.75	333.15	Von Hippel-Lindau disease gene
N_Geneseq_36:Q75974	+	34.00	96.75	867.54	pHLA-B7 expression vector. New
N_Geneseq_36:Q75973	+	34.00	94.99	1,1e+03	pHLA-B7/beta-2 microglobulin e
N_Geneseq_36:T96690	+	34.00	88.20	2,6e+03	Hereditary haemochromatosis 9
N_Geneseq_36:V57926	+	34.00	61.38	7,6e+04	Hereditary haemochromatosis
N_Geneseq_36:V57903	-	34.00	61.30	7,7e+04	Hereditary haemochromatosis
N_Geneseq_36:T93045	+	33.00	99.44	613.94	Mouse NF-AT Interacting Protei
N_Geneseq_36:X04354	+	33.00	98.89	659.23	Human secreted protein gene 44
N_Geneseq_36:Q29268	+	33.00	98.47	695.21	Human calcium channel 27980/10
N_Geneseq_36:Q08510	+	33.00	98.44	697.71	Sequence of anlyase gene and u
N_Geneseq_36:V59717	-	33.00	93.10	1,4e+03	Tumour rejection antigen precu
N_Geneseq_36:V59720	-	33.00	92.69	1,5e+03	Tumour rejection antigen precu
N_Geneseq_36:Q58522	-	33.00	90.69	1,9e+03	Sequence of Heliothis armigera
N_Geneseq_36:Q83735	+	33.00	88.39	2,5e+03	Calcium channel alpha-1E subun

N\_Geneseq\_36:Q84662 + 33.00 88.25 2.6e+03 7032 ! Human neuronal calcium chan  
N\_Geneseq\_36:Q84663 + 33.00 88.18 2.6e+03 7089 ! Human neuronal calcium chan  
N\_Geneseq\_36:Q86851 + 33.00 86.31 3.3e+03 8789 ! Human mitosis gene. Purifie  
N\_Geneseq\_36:V09076 + 33.00 86.31 3.3e+03 8789 ! Mitosis nucleic acid sequen

seq\_name: N\_Geneseq\_36:V30457

## seq\_documentation\_block:

ID V30457 standard; DNA; 1284 BP.

AC V30457:

DE 14-OCT-1998 (first entry)

DT Chimeric HLA-A2.1/beta-2 microglobulin coding sequence.

KW Antigen; major histocompatibility complex; MHC; lymphocyte; detection;

KW Immunisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;

KW Viral infection; chimeric; beta-2 microglobulin; ss.

OS Synthetic.

OS Homo sapiens.

FH Key

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

KW MHC; class I; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..1101  
 FT /\*tag= a  
 FT J03112485-A.  
 PN 14-MAY-1991.  
 PD 22-SEP-1989; 247695.  
 PF 22-SEP-1989; JP-247695.  
 PR (OLYU ) OLYMPUS OPTICAL KK.  
 PA WPI; 91-182989/25.  
 DR P-PSDB; R12465.  
 DT HLA-C gene, DNA probe and transformant cells - for immunisation  
 PT of animals and monoclonal antibody development.  
 PS Claim 1; Page 1; 13pp; Japanese.  
 CC Probes comprising part of the sequence can be used to identify  
 CC Class I genes. The DNA can be expressed for immunisation of  
 CC animals and prodn. of monoclonal antibodies specific for the HLA-C  
 CC antigen. See also Q12117 (same patent) and J03112486 and J03112487.  
 CC Sequence 1101 BP; 211 A; 337 C; 377 G; 176 T;  
 SQ

alignment\_scores:  
 Quality: 40.00 Length: 10  
 Ratio: 4.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 80.000

alignment\_block:

US-08-653-294-37 x Q12116 ..

Align seg 1/1 to: Q12116 from: 1 to: 1101

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
 |||||:|||||:|||||:|||||:|||||  
 295 CGAGTGAGCCTGCGGAACCTGCGGCTAC 324

seq\_name: N\_Geneseq\_36:Q12117

seq\_documentation\_block:

ID Q12117 standard; DNA; 1101 BP.  
 AC Q12117;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-C exon Cb-2.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..1101  
 FT /\*tag= a  
 FT J03112485-A.  
 PN 14-MAY-1991.  
 PD 22-SEP-1989; 247695.  
 PF 22-SEP-1989; JP-247695.  
 PR (OLYU ) OLYMPUS OPTICAL KK.  
 PA WPI; 91-182989/25.  
 DR P-PSDB; R12466.  
 DT HLA-C gene, DNA probe and transformant cells - for immunisation  
 PT of animals and monoclonal antibody development.  
 PS Claim 2; Page 1; 13pp; Japanese.  
 CC Probes comprising part of the sequence can be used to identify  
 CC Class I genes. The DNA can be expressed for immunisation of  
 CC animals and prodn. of monoclonal antibodies specific for the HLA-C  
 CC antigen. See also Q12116 (same patent) and J03112486 and J03112487.  
 CC Sequence 1101 BP; 215 A; 335 C; 379 G; 172 T;  
 SQ

alignment\_scores:  
 Quality: 40.00 Length: 10  
 Ratio: 4.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 80.000

alignment\_block:

US-08-653-294-37 x Q12117 ..

Align seg 1/1 to: Q12117 from: 1 to: 1101

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
 |||||:|||||:|||||:|||||:|||||  
 295 CGAGTGAGCCTGCGGAACCTGCGGCTAC 324

seq\_name: N\_Geneseq\_36:V21187

seq\_documentation\_block:

ID V21187 standard; DNA; 53789 BP.  
 AC V21187;  
 DT 24-JUL-1998 (first entry)  
 DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.  
 KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
 KW Amycolatopsis mediterranei; actinomycete; ansamycin; ds.  
 OS Amycolatopsis mediterranei.  
 FH Key Location/Qualifiers  
 FT CDS 1825..15543  
 FT /\*tag= a  
 FT /label= ORF\_A  
 FT /product= "polyketide synthase"  
 FT 15550..30759  
 FT /\*tag= b  
 FT /label= ORF\_B  
 FT /product= "polyketide synthase"  
 FT 30895..36050  
 FT /\*tag= c  
 FT /label= ORF\_C  
 FT /product= "polyketide synthase"  
 FT 36259..41325  
 FT /\*tag= d  
 FT /label= ORF\_D  
 FT /product= "polyketide synthase"  
 FT 41373..51614  
 FT /\*tag= e  
 FT /label= ORF\_E  
 FT /product= "polyketide synthase"  
 FT 51713..5293  
 FT /\*tag= f  
 FT /label= ORF\_F  
 FT /product= "polyketide synthase"  
 FT W09807868-A1.  
 PN 26-FEB-1998.  
 PD 18-AUG-1997; E04495.  
 PF 20-AUG-1996; EP-810551.  
 PR (NOVS ) NOVARTIS AG.  
 PA Engel N, Schupp T, Toupet C;  
 PI WPI; 98-169172/15.  
 DR P-PSDB; W52845-W52850.  
 DT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used  
 PT to produce rifamycin and rifamycin analogues  
 PS Claim 4; Page 53-102; 205pp; English.  
 CC The present sequence represents a Amycolatopsis mediterranei rifamycin  
 CC synthesis gene cluster DNA fragment from the present invention. The  
 CC DNA fragment comprises a DNA region involved directly or indirectly  
 CC in the gene cluster responsible for rifamycin synthesis, including  
 CC the adjacent DNA regions to the right and left which, by reason of  
 CC their function in connection with rifamycin biosynthesis, qualify  
 CC as constituents of this rifamycin gene cluster, and functional  
 CC fragments, derivatives or constituents of these. The Amycolatopsis  
 CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used  
 CC for producing rifamycin, rifamycin analogues or precursors. It can also  
 CC be used for inactivating or modifying genes involved in ansamycin or  
 CC rifamycin biosynthesis. The DNA can be used for constructing mutant  
 CC actinomycetes strains from which the natural rifamycin or ansamycin  
 CC biosynthesis gene cluster has been partly or completely deleted. The  
 CC DNA fragment can be used for assembling a library of polyketide  
 CC synthases, which can be used for assembling a library of polyketides.  
 CC A hybridisation probe of the invention can be used for identifying DNA  
 CC fragments involved in the biosynthesis of ansamycins.  
 CC Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T;  
 SQ



alignment\_scores:  
 Quality: 38.00 Length: 9  
 Ratio: 4.222 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 77.778

## alignment\_block:

US-08-653-294-37 x V21187 ..

Align seg 1/1 to: V21187 from: 1 to: 53789

1 ArgValAspLeuArgThrLeuArgGly 9

.....  
 20272 AAGTCGACCTGGCGGCGTGGCGGC 20298

seq\_name: N\_Geneseq\_36:V52305

## seq\_documentation\_block:

ID V52305 standard; DNA; 11770 BP.  
 AC V52305;  
 DT 23-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:172.  
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 OS Streptococcus pneumoniae.  
 PN WO9818931-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; U19588.  
 PR 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 PI Kunsch CA, Rosen CA;  
 DR WPI: 98-272225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1: Page 1094-1101; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 SQ Sequence 11770 BP; 3673 A; 1919 C; 2574 G; 3603 T;

## alignment\_scores:

Quality: 37.00 Length: 8  
 Ratio: 4.625 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 75.000

## alignment\_block:

US-08-653-294-37 x V52305/rev ..

Align seg 1/1 to reverse of: V52305 from: 1 to: 11770

3 AspLeuArgThrLeuArgGlyTyr 10  
 |||:|||||:|||||:|||||  
 11503 GATATGAGGACGTTAAAGGTTAC 11480

seq\_name: N\_Geneseq\_36:V52267

## seq\_documentation\_block:

ID V52267 standard; DNA; 12665 BP.  
 AC V52267;  
 DT 23-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:134.  
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 OS Streptococcus pneumoniae.  
 PN WO9818931-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; U19588.  
 PR 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 PI Kunsch CA, Rosen CA;  
 DR WPI: 98-272225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1: Page 908-915; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 SQ Sequence 12665 BP; 3902 A; 2431 C; 2890 G; 3441 T;

## alignment\_scores:

Quality: 37.00 Length: 8  
 Ratio: 4.625 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 75.000

## alignment\_block:

US-08-653-294-37 x V52267/rev ..

Align seg 1/1 to reverse of: V52267 from: 1 to: 12665

3 AspLeuArgThrLeuArgGlyTyr 10  
 |||:|||||:|||||:|||||  
 12528 GATATGAGGACGTTAAAGGTTAC 12505

seq\_name: N\_Geneseq\_36:T06769

## seq\_documentation\_block:

ID T06769 standard; DNA; 28598 BP.  
 AC T06769;  
 DT 15-OCT-1996 (first entry)  
 DE Sorangium cellulosum soraphen gene cluster.

```

alignment_scores:
  Quality: 37.00      Length: 10
  Ratio: 4.11       Gaps: 0
  Percent Similarity: 90.00    Percent Identity: 70.00

alignment_block:
  US-08-653-294-37 x T89956/rev ..
  Align seq 1/1 to reverse of: T89956 from: 1 to: 289558

```

1 ArgValAspLeuArgThrLeuArgGlyTyr 10  
3978 AGGGTCGATGCCCCGCACCGCTCGAAGGCTTC 3949

seq\_name: N\_Geneseq\_36:V05287

seq\_documentation\_block:  
ID: V05287 standard: DNA: 49377 BP:

DE The soraphen biosynthesis gene cluster from *Sorangium cellulosum*.  
 DT 21-MAY-1998 (first entry)  
 DE The soraphen biosynthesis gene cluster from *Sorangium cellulosum*.  
 KW Polyketide synthase; PKS; biosynthesis; soraphen; SorS; SorA; SorB;  
 KW SorD; biosynthetic module; beta-ketoacyl synthase; acyltransferase  
 KW ketoreductase; beta-ketone processing domain; cytosolic agent;  
 KW antimicrobial agent; phytopathogenic fungi; transgenic plant;  
 KW biological control; ss.

OS	OS program	Location/Qualifiers
FH	Key	

```

FI      /product= sorb
FT      /note= "gene product highly homologous to the
FT      reductase domains of type I PKss such
FT      as ervA from saccharopolyspora erythraea"
FT

```

FT CDS 927. .19874

FT /\*tag= b

FT /note= "gene product is highly homologous to

FT type I PKSS that are known to be involved

FT — in the synthesis of polyketide compounds"

```
FT      misc_feature 942.7115
      /4425.0
```

```

E# /tag= c
E# /note= "module 1 of sorA"

```

FT	misc feature	7203.12884	/note= module 1 of 5014
----	--------------	------------	-------------------------

```

FI      / *tag= d
FIF     / *tag= d

```

```
/note= "module 2 of SorA"
```

```
FT misc_feature 13455. 19616
```

FT /\*tag= e

```
FT      /note= "module 3 of SorA"
```

FT	misc_feature	19871.	.46318
nm		/4422=	5

10871 16318

```

CDS
198/1: .40318
/*tag= q

```

```

LT          /csg= 9
FT          /product= sorB

```

FT  
I  
/note= "gene product is highly homologous to

FT  
type I PKS genes"

```
FT misc_feature 19870. 24556
```

```

IT      h
/*tag=

```

```

EFT 20820
/NOTE= "module 1 of sor8"
24538 20820

```

FT  
misc\_reature  
24638: .30820  
/tag= 1  
PM

```

Z1      /:tag=
FT      /note= "module 2 of sorB"

```

FT	misc_feature	30881.	.35446
----	--------------	--------	--------

```

ET
-----
/*tag= j

```

```
FT /note= "module 3 of SorB"
```

FT	misc_feature	35528.	.40114
FT			

```

EE k
EE /*tag=
EE "module A of sorb"

```

FT	misc feature	40190	46318	/note=	module 4 of solb
FT					

```

#0130: 1.40310
/*tag= 1

```

```
FT
/
/note= "module 5 of sorB"
```



US-08-653-294-37 x X23032/rev ..

Align seg 1/1 to reverse of: X23032 from: 1 to: 7207

1 ArgValAspLeuArgThrLeuArgGly 9

|||||  
1536 CCGGTGCGACCTCAGAACCTCATAGGC 1510

seq\_name: N\_Geneseq\_36:X20248\_06

seq\_documentation\_block:

Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi polynucleotide s

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

Fragment Name	Begin	End
WP X20248_00	1	110000
WP X20248_01	100001	210000
WP X20248_02	200001	310000
WP X20248_03	300001	410000
WP X20248_04	400001	510000
WP X20248_05	500001	610000
WP X20248_06	600001	710000
WP X20248_07	700001	810000
WP X20248_08	800001	910000
WP X20248_09	900001	910715

alignment\_scores:  
Quality: 36.00 Length: 10  
Ratio: 4.000 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 60.000

alignment\_block:  
US-08-653-294-37 x X20248\_06/rev ..

Align seg 1/1 to reverse of: X20248\_06 from: 1 to: 110000

1 ArgValAspLeuArgThrLeuArgGlyTyr 10

|||||  
101283 AGAATTGACATAGGAACTCTTGAGGGTTAT 101254

seq\_name: N\_Geneseq\_36:X20248\_07

seq\_documentation\_block:

Continuation (8 of 10) of X20248 from base 700001 (Borrelia burgdorferi polynucleotide s

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

Fragment Name	Begin	End
WP X20248_00	1	110000
WP X20248_01	100001	210000
WP X20248_02	200001	310000
WP X20248_03	300001	410000
WP X20248_04	400001	510000
WP X20248_05	500001	610000
WP X20248_06	600001	710000
WP X20248_07	700001	810000
WP X20248_08	800001	910000
WP X20248_09	900001	910715

alignment\_scores:  
Quality: 36.00 Length: 10  
Ratio: 4.000 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 60.000

alignment\_block:  
US-08-653-294-37 x X20248\_07/rev ..

Align seg 1/1 to reverse of: X20248\_07 from: 1 to: 110000

1 ArgValAspLeuArgThrLeuArgGlyTyr 10

|||||  
1283 AGAATTGACATAGGAACTCTTGAGGGTTAT 1254

seq\_name: N\_Geneseq\_36:T17935

seq\_documentation\_block:

ID T17935 standard; CDNA; 360 BP.

AC T17935;

DT 16-AUG-1996 (first entry)

DE Taenia ovis antigenic polypeptide coding sequence.

KW Taenia ovis; antigen; vaccine; infection; cestode; parasite; probe;

OS viral vaccine; Echinococcus; Taenid parasite; ss.

FT Key

FT cds Location/Qualifiers

1..360

/tag= a

/product= Antigenic polypeptide.

PN ZA9402410-A.

PD 31-JAN-1996.

PF 07-APR-1994; 002410.

PR 07-APR-1993; NZ-237361.

PA (PITM ) PITMAN MOORE NZ LTD.

PI Dempster RP, Gauci C, Harrison GBL, Heath DD, Lawrence SB;

PI Lightowlers MW, Richard MD, Robinson CM;

DR WPI; 96-139972/14.

DR F-PSDB; R92832.

PT Taenia ovis antigenic polypeptide - used to develop vaccines to

PT protect against infection by a cestode parasite, partic. in

PT ruminants.

PS Claim 11: Page 22-24; 36pp; English.

CC The antigenic polypeptide encoded by this sequence, its fragments

CC and variants can be used in a vaccine to protect against infection

CC by a cestode parasite, particularly in ruminants. The nucleic acids

CC (See also T17934) can be used to produce recombinant viral vaccines.

CC The DNA (preferably this DNA sequence) can also be labelled and used

CC as a probe to identify nucleic acids encoding a protective antigen

CC of an Echinococcus or Taenid parasite other than Taenia ovis.

CC Sequence 360 BP; 98 A; 99 C; 86 G; 77 T;

SQ

alignment\_scores:

Quality: 35.00 Length: 8

Ratio: 4.375 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:

US-08-653-294-37 x T17935/rev ..

Align seg 1/1 to reverse of: T17935 from: 1 to: 360

1 ArgValAspLeuArgThrLeuArg 8

|||||  
175 AGACTGACTTGGACACTGAGA 152

seq\_name: N\_Geneseq\_36:Q72739

seq\_documentation\_block:

ID Q72739 standard; CDNA; 600 BP.

AC Q72739;

DT 10-MAY-1995 (first entry)

DE T. ovis vaccine candidate antigen.

KW Vaccine, protective antigen; tapeworm; glutathione-S-transferase;

KW fusion protein; Escherichia coli; pGEX-21; ss.

OS Taenia ovis.

FT Key Location/Qualifiers

10..372

/tag= a

FT WO9422913-A.

PD 13-OCT-1994.

PF 07-APR-1994; NZ0029.

PR 07-APR-1993; NZ-247361.

PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.

PA (PITM ) PITMAN MOORE NZ LTD.

PA (UYNE ) UNIV MELBOURNE.

PI Dempster RP, Gauci C, Harrison GBL, Heath DD, Lawrence SB;

PI Lightowlers MW, Richard MD, Robinson CM;

DR WPI; 94-333117/41.

DR P-PSDB; RG2045.  
PT New protective antigen from Taenia ovis - and related DNA  
PT vectors, transformed cells and antibodies, useful in vaccines,  
PT also for detecting homologous genes and proteins  
PS Disclosure: Page 22-23; 37pp; English.  
CC A T. ovis oncosphere lambda gt11 cDNA library was screened with  
CC rabbit antibodies raised against a 16 kDa protective antigen of T.  
CC ovis. The insert from a selected clone was subcloned into vector  
CC pGEX-2T and expressed as a glutathione-S-transferase fusion protein  
CC in Escherichia coli JM101. The DNA and predicted amino acid  
CC sequences of the T. ovis portion of the fusion protein were  
CC determined.  
SQ Sequence 600 BP; 194 A; 153 C; 121 G; 132 T;

alignment\_scores:  
Quality: 35.00 Length: 8  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:

US-08-653-294-37 x Q72739/rev ..

Align seg 1/1 to reverse of: Q72739 from: 1 to: 600

1 ArgValAspLeuArgThrLeuArg 8  
||||:|||||||  
184 AGACTGGACTTGGACACTGAGA 161

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-37 to: EST:\* out\_format : pfs

Date: Feb 8, 2000 6:23 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+ p2n model -DEV=xlp  
-Q/cnrl\_1/USPTO.spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.2  
-DB-EST -QFMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPEXT=0.500 -DELOP=6.000  
-DLEFT=7.000 -START=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pt -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPUP=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-37

Query length: 10

Database: EST:\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 7600.090000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est4:H35352	+	51.00	0.1390	351	H35352 EST109782 Rat PC-12 cell
gb_est13:AA376928	+	45.00	195.80	113	AA376928 EST89412 Small intestine
gb_est13:AA377133	+	46.00	0.4272	128	AA377133 EST89666 Small intestine
gb_est36:AA900799	+	46.00	175.31	338	AA900799 UI-R-E0-dn-h-09-0-UI-S
gb_est3:C17443	+	46.00	187.02	356	C17443 C17443 Human placenta ct
gb_est12:AA313847	+	46.00	166.55	377	AA313847 EST185704 Colon carcin
gb_est11:AA243092	+	46.00	166.02	420	AA243092 zr25b01.r1 Stratagene
gb_est32:AA1750810	+	46.00	165.04	435	AA1750810 cn05g02.x1 Normal Hum
gb_est10:AA176192	+	46.00	164.71	448	AA176192 zp24g07.r1 Stratagene
gb_est34:AA1214149	+	46.00	164.45	469	AA1214149 ap30e06.x1 Schiller as
gb_est36:AA1214126	+	46.00	163.42	501	AA1214126 ap30c06.x1 Schiller as
gb_est36:AA102538	+	46.00	161.07	648	AA102538 EST211827 Normalized i
gb_est13:AA313054	+	44.00	162.32	245	AA313054 EST183851 Pancreas tum
gb_est23:AA1158411	+	44.00	159.21	344	AA1158411 ud24e07.r1 Soares 2ND
gb_est3:AA125323	-	44.00	158.25	382	AA125323 mp80c08.r1 Soares 2ND
gb_est10:AA134053	+	44.00	158.23	383	AA134053 mp80c08.r1 Soares 2ND
gb_est30:AA166560	+	44.00	157.99	393	AA166560 mul4d08.x1 Soares 2ND
gb_est19:AA764120	+	44.00	157.90	397	AA764120 vv45b03.r1 Soares 2ND
gb_est2:AA12066	+	44.00	157.72	405	AA12066 vf54d10.r1 Soares infant
gb_est10:AA132653	+	44.00	157.21	428	AA132653 z021a05.r1 Stratagene
gb_est25:AA1316043	+	44.00	157.04	436	AA1316043 u161f12.y1 Sugano mous
gb_est11:AA193357	+	44.00	157.00	438	AA193357 mul4n06.r1 Soares 2ND
gb_est19:AA790185	+	44.00	156.88	444	AA790185 vv82e04.r1 Stratagene
gb_est12:AA305941	+	44.00	156.74	451	AA305941 EST176934 Jurkat T-cel
gb_est10:AA175455	+	44.00	156.63	456	AA175455 ms87h04.r1 Soares mous
gb_est20:AA880833	+	44.00	156.63	456	AA880833 vx46d06.r1 Stratagene
gb_est8:AA017870	-	44.00	156.59	458	AA017870 mh47h06.r1 Soares mous
gb_est4:H23377	+	44.00	156.57	459	H23377 ym57e02.r1 Soares infant
gb_est3:AA028838	-	44.00	156.53	461	AA028838 mh90f04.r1 Soares mous
gb_est14:AA396998	-	44.00	156.53	461	AA396998 mx86e07.r1 Soares mous
gb_est34:AA1787144	+	44.00	156.52	462	AA1787144 u185d06.y1 Sugano mous
gb_est15:AA488534	+	44.00	156.22	477	AA488534 ac37f08.r1 Stratagene
gb_est26:AA136563	+	44.00	156.11	478	AA136563 ac90e01.x1 Schiller me
gb_est37:AA012806	+	44.00	155.31	527	AA012806 uc03a02.y1 Sugano mous
gb_est17:AA604923	+	44.00	155.21	533	AA604923 nc81b07.s1 NCI CGAP AA
gb_est21:AA986072	+	44.00	155.17	536	AA986072 uc72f11.y1 Sugano mous
gb_est16:AA548636	+	44.00	155.10	539	AA548636 nj38d02.s1 NCI CGAP AA
gb_est23:AA116479	+	44.00	154.90	551	AA116479 ud74c02.y1 Sugano mous
gb_est35:AA1286941	+	44.00	154.15	598	AA1286941 u179a11.y1 Sugano mous
gb_est36:AA1876777	+	44.00	153.90	615	AA1876777 u135f09.y1 Sugano mous
gb_est34:AA1785805	+	44.00	152.53	714	AA1785805 u178g02.y1 Sugano mous

gb\_est11:AA197564 - 44.00 152.37 10.12 727 ! AA197564 mu22d05.r1 Soares 2  
gb\_est28:AA529689 + 44.00 152.03 10.56 754 ! AA529689 u181c04.y1 Sugano m  
gb\_est22:AA047292 + 44.00 151.31 11.59 816 ! AA047292 ud66e05.y1 Sugano m

seq\_name: gb\_est4:H35352

seq\_documentation\_block:

LOCUS H35352 351 bp mRNA EST 02-APR-1998  
DEFINITION EST109782 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. CDNA  
clone RPNAU72 similar to MHC class I, mRNA sequence.

ACCESSION H35352.1 GI:980769

VERSION H35352

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 351)

AUTHORS Lee, N.H., Weinstock, K.G., Kirkness, E.F., Earle-Hughes, J.A.,

Fuldner, R.A., Marmaras, S., Glodek, A., Gocayne, J.D., Adams, M.D.,

Kerlavage, A.R., Fraser, C.M. and Venter, J.C.

Comparative expressed-sequence-tag analysis of differential gene  
expression profiles in PC-12 cells before and after nerve growth  
factor treatment

Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)

JOURNAL 95396786

MEDLINE 95396786

COMMENT On May 8, 1995 this sequence version replaced gi:799766.

Other ESTs: TC508

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

For clone availability please contact the TIGR Database  
(tdbinfo@tigr.org) TC (tentative Consensus) numbers represent  
assemblies of ESTs.

FEATURES

source

1..351

/organism="Rattus sp."

/db\_xref="ATCC (inhost):2004068"

/db\_xref="taxon:10118"

/clone="RPNAU72"

/clone.lib="Rat PC-12 cells, NGF-treated (9 days)"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; poly(A)+ RNA was purified from 9-day NGF treated

PC12 cells. cDNA was constructed using an oligo-dT primer

and directionally cloned using the Lambda ZAP II Vector

Kit by Stratagene"

BASE COUNT 84 a 86 c 125 g 53 t 3 others

ORIGIN

alignment\_scores:

Quality: 51.00 Length: 10

Ratio: 5.100 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-653-294-37 x H35352

Align seg 1/1 to: H35352 from: 1 to: 351

1 ArgValAspLeuAthrLeuArgGlyTyr 10

|||||

|||||

37 CGAGTGGACCTGAGACCTCGCGGCTAC 66

seq\_name: gb\_est13:AA376928

seq\_documentation\_block:

LOCUS AA376928 113 bp mRNA EST

DEFINITION EST89412 Small intestine I Homo sapiens CDNA 5' end similar to

21-APR-1997





```

seq_name: gb_est26:AA900799

seq_documentation_block:
  338 bp      mRNA      EST      05-FEB-1999
LOCUS      AA900799
DEFINITION      UI-R-E0-dn-h-09-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
KEYWORDS      UI-R-E0-dn-h-09-0-UI 3', similar to gi11263201|gb|U38972|RN038972
SOURCE      Rattus norvegicus MHC class I RT1.Au heavy chain precursor, mRNA,
complete cds, mRNA sequence.
ACCESSION      AA900799
VERSION      AA900799.1 GI:4233294
KEYWORDS      EST.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 338)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      On Apr 7, 1998 this sequence version replaced gi:3036153.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dr track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics
Seq primer: M13 Forward.
FEATURES
  Location/Qualifiers
    1..338
     /organism="Rattus norvegicus"
     /strain="Sprague-Dawley"
     /db_xref="taxon:10116"
     /clone="UI-R-E0-dn-h-09-0-UI"
     /dev_stage="embryonic"
     /lab_host="DH10B (Life Technologies)"
     /note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; this library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT      62 a 104 c 124 g 47 t 1 others
ORIGIN
alignment_scores:
  Quality: 46.00      Length: 10
  Ratio: 4.600      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 90.000
alignment_block:
  US-08-653-294-37 x AA900799
  Align seg 1/1 to: AA900799 from: 1 to: 338
  1 ArgValAspLeuArgThrLeuArgGlyTyr 10
  |||||:|||||:|||||:|||||:|||||:
  255 CGAGTGAACCTGAGGACCTCGCGGCTAC 284

seq_name: gb_est9:C17443
seq_documentation_block:
  377 bp      mRNA      EST      19-APR-1997
LOCUS      AA313847
DEFINITION      EST185704 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5'
end similar to similar to major histocompatibility complex, class
I, E (GB:M20022), mRNA sequence.
ACCESSION      AA313847
VERSION      AA313847.1 GI:1966176
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 377)
AUTHORS      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palmarques,R.F., McDonald,L.A., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
seq_documentation_block:
  377 bp      mRNA      EST      19-APR-1997
LOCUS      AA313847
DEFINITION      EST185704 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5'
end similar to similar to major histocompatibility complex, class
I, E (GB:M20022), mRNA sequence.
ACCESSION      AA313847
VERSION      AA313847.1 GI:1966176
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 377)
AUTHORS      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palmarques,R.F., McDonald,L.A., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

```

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 On Dec 3, 1996 this sequence version replaced gi:975987.  
 Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1647 Std Error: 0.00  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 356.

```

FEATURES
    source
        1. .420
            /organism="Homo sapiens"
            /db_xref="GDB:5426385"
            /db_xref="taxon:9606"
            /clone="IMAGE:564393"
            /clone_lib="Stratagene NT2 neuronal precursor 937230"
            /tissue.type="neuroepithelial cells"
            /dev.stage="Ntera-2 neuroepithelial cells"

```

```

/seq_name="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-GAP XR vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG
3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
72 a      134 c      139 g      75 t
BASE COUNT
ORIGIN

alignment_scores:
Quality:       46.00
Length:        10

```

```

alignment_block:
  US-08-653-294-37 x AA243092 ..
  Align seg 1/1 to: AA243092 from: 1 to: 420

  1 ArgValAspLeuArgThrLeuArgGlyTyr 10
  |||||:|||||:|||||:|||||:|||||
  284 CGAGTGAACCTGGGACGCTGCGCGCTAC 313

seg_name: gb_est32:AI750810

seg_documentation_block:
  LOCUS AI750810 435 bp mRNA
  DEFINITION clone n05g02.xl Normal Human Trabecular Bone Cells Homo sapiens cdna
  clone NHTBC cn05d02 random, mRNA sequence.
  EST 22-JUN-1999

```

VERSION	01.01.94
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.
	1 (bases 1 to 435)
	Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,
	Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M.,
	Robay,P.G., Hotchkiss,R.N. and Francomano,C.A.
TITLE	SGAP: The Skeletal Genome Anatomy Project
JOURNAL	Unpublished (1997)
COMMENT	On Feb 18, 1999 this sequence version replaced gi:4296314.
	Contact: Libin Jia
	Medical Genetics Branch
	National Human Genome Research Institute



```

/sex="male"
/tissue_type="astrocytoma"
/dev_stage="44 years"
/lab_host="SOLR"
/Note="Organ: brain; Vector: pBluescript SK- (Stratagene);
Site_1: EcoRI; Site_2: XhoI; Double-stranded cDNA was
prepared from human astrocytoma using primer
5'-GAGAGAGAGAGAGAGAACTAGTCTGAGT(18)-3'. An EcoRI
adaptor was used on the 5' end of the cDNA as follows:
5'-AATTCGACGAG-3'. The library was size-selected and
went through one round of amplification. Average insert
size is 1.7 kb, with a range from 0.4-12 kb. Tumor
identification by consensus pathology. This library was
constructed by Dr. Martin Schiller (Johns Hopkins
University)."
BASE COUNT      93 a 141 c 157 g 78 t
ORIGIN

alignment_scores:
  Quality: 46.00      Length: 10
  Ratio: 4.600       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 90.000

alignment_block:
US-08-653-294-37 x AI214149 ..
Align seg 1/1 to: AI214149 from: 1 to: 469
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||:|||||:|||||:|||||:|||||
199 CGAGTGAACCTGGGACGCTGCGGGCTAC 228

seq_name: gb_est24:AI214126

seq_documentation_block:
LOCUS      AI214126      501 bp      mRNA      21-OCT-1998
DEFINITION      ap30c06.x1 Schiller astrocytoma Homo sapiens cDNA clone
IMAGE:1956874 3' similar to gb:M20022 HLA CLASS I
HISTOCOMPATIBILITY ANTIGEN, E*0101/E*0102 ALPHA (HUMAN);, mRNA
sequence.
ACCESSION      AI214126
VERSION        AI214126.1 GI:3777727
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 501)
AUTHORS        Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,T., Wyllie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
JOURNAL
COMMENT        On Jan 17, 1998 this sequence version replaced gi:1900329.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 295.
Location/Qualifiers
1..501
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1956874"
/clone_lib="Schiller astrocytoma"
/sex="male"

```

```

/tissue_type="astrocytoma"
/dev_stage="44 years"
/lab_host="SOLR"
/Note="Organ: brain; Vector: pBluescript SK- (Stratagene);
Site_1: EcoRI; Site_2: XhoI; Double-stranded cDNA was
prepared from human astrocytoma using primer
5'-GAGAGAGAGAGAGAGAACTAGTCTGAGT(18)-3'. An EcoRI
adaptor was used on the 5' end of the cDNA as follows:
5'-AATTCGACGAG-3'. The library was size-selected and
went through one round of amplification. Average insert
size is 1.7 kb, with a range from 0.4-12 kb. Tumor
identification by consensus pathology. This library was
constructed by Dr. Martin Schiller (Johns Hopkins
University)."
BASE COUNT      99 a 148 c 170 g 81 t
ORIGIN

alignment_scores:
  Quality: 46.00      Length: 10
  Ratio: 4.600       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 90.000

alignment_block:
US-08-653-294-37 x AI214126 ..
Align seg 1/1 to: AI214126 from: 1 to: 501
1 ArgValAspLeuArgThrLeuArgGlyTyr 10
|||||:|||||:|||||:|||||:|||||
199 CGAGTGAACCTGGGACGCTGCGGGCTAC 228

seq_name: gb_est26:AI102538

seq_documentation_block:
LOCUS      AI102538      648 bp      mRNA      08-JAN-1999
DEFINITION      EST211827 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMOB096 3' end, mRNA sequence.
ACCESSION      AI102538
VERSION        AI102538.1 GI:4134091
KEYWORDS       EST.
SOURCE         Rattus sp.
ORGANISM       Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 648)
AUTHORS        Lee,N.H., Glodet,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
Unpublished (1998)
COMMENT        On Aug 21, 1998 this sequence version replaced gi:3704683.
Other ESTs: TC51871
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..648
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone_lib="REMOB096"
/dev_stage="embryo 8, 12, 18 dpc"
/Note="vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      129 a 189 c 228 g 102 t
ORIGIN

```



## ORIGIN

BASE COUNT 55 a 124 c 107 g 96 t  
ORIGIN

## alignment\_scores:

Quality: 44.00 Length: 10  
Ratio: 4.889 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 90.000

## alignment\_block:

US-08-653-294-37 x A1158411/rev ..

Align seg 1/1 to reverse of: A1158411 from: 1 to: 344

1 ArgValAspLeuArgThrLeuArgGlyTyr 10

|||||  
141 CGAGTGGACCTGAGGACCTGCTCGGCTAC 112

seq\_name: gb\_est9:A1125323

## seq\_documentation\_block:

LOCUS A1125323 382 bp mRNA EST 18-FEB-1997  
DEFINITION mp80c08.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:575534 5' similar to gb:X00492\_cds1 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-3 A\*0301 ALPHA (HUMAN); gb:M69073 Mus musculus mRNA, complete cds (MOUSE);, mRNA sequence.

## ACCESSION

A1125323

## VERSION

A1125323.1 GI:1684491

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 382)

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced gi:1282198.  
Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:350182

Seq primer: -28M13 rev2 from Amersham.

## FEATURES

source

1..382  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:575534"  
/clone\_lib="Soares 2NbMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
note=vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, TGTACCAATCTGAAGTGGAGCGCGCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

alignment\_scores:  
Quality: 44.00 Length: 10  
Ratio: 4.889 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 90.000  
alignment\_block:  
US-08-653-294-37 x A1125323/rev ..

Align seg 1/1 to reverse of: A1125323 from: 1 to: 382

1 ArgValAspLeuArgThrLeuArgGlyTyr 10

|||||  
344 CGAGTGGACCTGAGGACCTGCTCGGCTAC 315

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 13:17:41 ; Search time 209.03 seconds  
(without alignments)  
3.317 Million cell updates/sec

Title: US-08-653-294-15  
Perfect score: 49  
Sequence: 1 YRLIIRLDER 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

SPTREMBL\_12:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	75.5	151	038407	O58407 pyrococcus
2	35	71.4	506	11	Q62157 mus musculus
3	34	69.4	87	2	Q33348 mycobacteri
4	33	67.3	132	2	Q55457 synechocyst
5	33	67.3	469	2	O83484 treponema p
6	33	67.3	577	10	Q9XJ30 oryza sativ
7	33	67.3	1324	2	Q44103 Q41032 amycolotops
8	32	65.3	52	8	O20152 chlorella v
9	32	65.3	267	5	O9Y058 caenorhabdi
10	32	65.3	317	4	O95006 homo sapien
11	32	65.3	341	5	Q25032 haemochus
12	32	65.3	342	3	Q02895 saccharomyc
13	32	65.3	428	2	Q9X314 bacillus an
14	32	65.3	441	5	O16954 caenorhabdi
15	32	65.3	554	2	O06421 mycobacteri
16	32	65.3	759	5	O61660 caenorhabdi
17	32	65.3	982	4	O34942 homo sapien
18	32	65.3	1393	4	O75872 homo sapien
19	32	65.3	1400	2	P96419 mycobacteri
20	32	65.3	1440	2	O69498 mycobacteri

21	31	63.3	284	2	O05385
22	31	63.3	290	5	O9XV35
23	31	63.3	312	2	O9X7V5
24	31	63.3	333	2	O54274
25	31	63.3	335	8	O98296
26	31	63.3	395	10	O9ZT98
27	31	63.3	418	2	O9X8L1
28	31	63.3	495	5	O94013
29	31	63.3	515	2	O9ZAP8
30	31	63.3	560	13	O08781
31	31	63.3	1808	13	O42142
32	30	61.2	55	10	O41314
33	30	61.2	123	10	O41313
34	30	61.2	149	3	O9Y703
35	30	61.2	155	6	O97931
36	30	61.2	156	2	P71966
37	30	61.2	173	2	Q51316
38	30	61.2	182	13	O9W661
39	30	61.2	182	13	O9W659
40	30	61.2	198	2	P96676
41	30	61.2	235	1	O57883
42	30	61.2	277	2	Q9ZB18
43	30	61.2	287	2	Q44214
44	30	61.2	297	2	O32477
45	30	61.2	303	2	O69350

## ALIGNMENTS

RESULT 1

O58407  
ID O58407 PRELIMINARY; PRT; 151 AA.  
AC O58407;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
DE 151AA LONG HYPOTHETICAL FRXA PROTEIN.  
GN PH0674.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE: 98344137.  
RA KAWABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAMA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
RA KIKUCHI H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL; AP000003; BAA29765.1;  
SQ SEQUENCE 151 AA; 17160 MW; 11AACD59 CRC32;

Query Match 75.5%; Score 37; DB 1; Length 151;  
Best Local Similarity 88.9%; Pred. No. 4.8;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 RLLIIRLDER 10  
|||||  
Db 96 RLLIIRLDER 104

RESULT 2

O62157 PRELIMINARY; PRT; 506 AA.  
ID O62157  
AC O62157;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
 DE ZINC FINGER PROTEIN (FRAGMENT).  
 GN RFP.  
 OC Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RA TAKAHASHI M.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.  
 DR EMBL; X75343; CAA53092.1; -;  
 DR PROSITE; PS00518; ZINC\_FINGER\_C3HC4; 1.  
 DR PFAM; PF00622; SPRY; 1.  
 DR PFAM; PF00643; zf-B\_box; 1.  
 DR PFAM; PF00097; zf-C3HC4; 1.  
 DR DNA-binding; Zinc-finger.  
 FT NON\_TER 1  
 SQ SEQUENCE 506 AA; 57882 MW; AEE397C3 CRC32;

Query Match 71.4%; Score 35; DB 11; Length 506;  
 Best Local Similarity 77.8%; Pred. No. 40;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLRLRDE 9  
 ||||| 1:1  
 Db 191 YRLRLRLEE 199

RESULT 3  
 ID O33348 PRELIMINARY; PRT; 87 AA.  
 AC O33348;  
 DT 01-JAN-1998 (TReMBLrel. 05, Created)  
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
 DE HYPOTHETICAL 10.2 KD PROTEIN.  
 GN MYV003.12.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA MURPHY L., HARRIS D.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA MEDLINE; 96181548.  
 RA PHILIPP W.J., FOULET S., EIGLMEIER K., PASCOPELLA L.,  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RL "An integrated map of the genome of the tubercle bacillus,  
 RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium  
 RT leprae".  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).  
 DR EMBL; AL008883; CAA15528.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 87 AA; 10237 MW; 62B840DE CRC32;

Query Match 69.4%; Score 34; DB 2; Length 87;  
 Best Local Similarity 66.7%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLRLRDE 9  
 ||||| 1:1  
 Db 60 YRLRLRIDD 68  
 RESULT 4  
 ID Q55457 PRELIMINARY; PRT; 132 AA.  
 AC Q55457;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)  
 DE HYPOTHETICAL 15.2 KD PROTEIN.  
 OS Synecocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA TABATA S.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RX MEDLINE; 96127529.  
 RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,  
 RA SUGIURA M., TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RX MEDLINE; 97061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M.,  
 RA TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 KW EMBL; D64006; BAA10800.1; -;  
 SQ SEQUENCE 132 AA; 15240 MW; 53B23658 CRC32;

Query Match 67.3%; Score 33; DB 2; Length 132;  
 Best Local Similarity 66.7%; Pred. No. 27;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLRLRDE 9  
 ||||| 1:1  
 Db 120 YRLRLRIDD 128

RESULT 5  
 ID O83484 PRELIMINARY; PRT; 469 AA.  
 AC O83484;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
 DE HYPOTHETICAL 53.0 KD PROTEIN.  
 GN TP0471.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98332770.  
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,



RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,  
 RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT Spirochete";  
 RL Science 281:375-388(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
 RA DODSON R., GWINN M., HICKY E.K., CLAYTON R., KETCHUM K.A.,  
 RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,  
 RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE001223; AAC65460.1; -;  
 DR TIGR; TP0471; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 469 AA; 52960 MW; F76800E7 CRC32;

Query Match 67.3%; Score 33; DB 2; Length 469;  
 Best Local Similarity 87.5%; Pred. No. 94;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLRLD 8  
 Db 140 YTLRLD 147

RESULT 6  
 Q9XJ30 PRELIMINARY; PRT; 577 AA.  
 AC Q9XJ30;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE ANTHRANILATE SYNTHASE ALPHA 1 SUBUNIT.  
 GN OSASAL.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC eukaryotes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 OC Poaceae; Oryza.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA TOZAWA Y., HASEGAWA H., TERAKAWA T., WAKASA K.;  
 RT "Rice cDNA encoding anthranilate synthase alpha subunit.";  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB022602; BAA82094.1; -;  
 SQ SEQUENCE 577 AA; 63947 MW; 057FE09C CRC32;

Query Match 67.3%; Score 33; DB 10; Length 577;  
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLRLD 10  
 Db 81 YRCLVRED 90

RESULT 7  
 Q44103 PRELIMINARY; PRT; 1324 AA.  
 AC Q44103;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE PEPTIDE-SYNTHETASE (FRAGMENT).

GN APS.  
 OS Amycolatopsis mediterranei.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiales; Amycolatopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 5908;  
 RX MEDLINE; 97449857.  
 RA PELZER S., REICHERT W., HUPPERT M., HECKMANN D., WOHLLEBEN W.;  
 RT "Cloning and analysis of a peptide synthetase gene of the ballmucin  
 RT producer Amycolatopsis mediterranei DSM5908 and development of a gene  
 RT disruption/replacement system.";  
 RL J. Biotechnol. 56:115-128(1997).  
 DR EMBL; X97860; CAA66454.1; -;  
 DR PROSITE; PS00455; AMP-BINDING; 1.  
 DR PFAM; PF00501; AMP-BINDING; 1.  
 DR PFAM; PF00668; DUF4; 2.  
 DR PFAM; PF00550; pp-binding; 1.  
 KW Ligase.  
 FT NON\_TER 1  
 FT NON\_TER 1324 1324  
 SQ SEQUENCE 1324 AA; 142666 MW; 2C08588E CRC32;

Query Match 67.3%; Score 33; DB 2; Length 1324;  
 Best Local Similarity 70.0%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLRLD 10  
 Db 969 YRVAGRLD 978

RESULT 8  
 O20152 PRELIMINARY; PRT; 52 AA.  
 AC O20152;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE ORF52B.  
 OS Chlorella vulgaris.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;  
 OC Chlorellaceae; Chlorella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97303241.  
 RA WAKASUGI T., NAGAI T., KAPOOR M., SUGITA M., ITO M., ITO S.,  
 RA TSUDZUKI J., NAKASHIMA K., TSUDZUKI T., SUZUKI Y., HAMADA A., OHTA T.,  
 RA INAMURA A., YOSHINAGA K., SUGIURA M.;  
 RT "Complete nucleotide sequence of the chloroplast genome from the green  
 RT alga Chlorella vulgaris: the existence of genes possibly involved in  
 RT chloroplast division.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
 DR EMBL; AB001684; BAA20659.1; -;  
 KW Chloroplast.  
 SQ SEQUENCE 52 AA; 5985 MW; 74BE6B74 CRC32;

Query Match 65.3%; Score 32; DB 8; Length 52;  
 Best Local Similarity 60.0%; Pred. No. 17;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLRLD 10  
 Db 17 FLLVLD 26

RESULT 9  
 Q9Y058 PRELIMINARY; PRT; 267 AA.  
 ID Q9Y058  
 AC Q9Y058;

DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE RADI-LIKE PROTEIN.  
 GN HPR-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Secernentea; Rhabditia; Rhabditida;  
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderidae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 99097342.  
 RA DEAN F.B., LIAN L., O'DONNELL M.;  
 RT "CDNA cloning and gene mapping of human homologs for  
 RT Schistosaccharomyces pombe rad17, rad1, and hus1 and cloning of  
 RT homologs from mouse, Caenorhabditis elegans, and Drosophila  
 RT melanogaster.";  
 RL Genomics 54:424-436(1998).  
 DR EMBL: AF076843; AAC95525.1; -.  
 SQ SEQUENCE 267 AA; 29910 MW; A91CD2BF CRC32;

Query Match 65.3%; Score 32; DB 5; Length 267;  
 Best Local Similarity 55.6%; Pred. No. 86;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLRLRLDER 10  
 |||||  
 DB 229 KLLRVDER 237

RESULT 10  
 O95006 PRELIMINARY; PRT; 317 AA.  
 AC O95006;  
 DT 01-MAY-1999 (TRENBLrel. 10, Created)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
 DE WUGSC:H.DJ0669B10.1 PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BAUER C., WILLIAMS D.;  
 RT "The sequence of Homo sapiens PAC clone DJ0669B10.";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.H.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004853; AAC64378.1; -.  
 SQ SEQUENCE 317 AA; 35293 MW; B32AE3AE CRC32;

Query Match 65.3%; Score 32; DB 4; Length 317;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLRLRLDER 10  
 |||||  
 DB 47 LLRLRLDER 54

RESULT 11  
 Q25032 PRELIMINARY; PRT; 341 AA.  
 ID Q25032  
 AC Q25032;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CYSTEINE PROTEINASE (FRAGMENT).  
 OS Haemochus contortus.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Strongylida;  
 OC Trichostrongyloidea; Trichostrongylidae; Haemonchinae; Haemonchus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92244291.  
 RA PRATT D., ARMES L.G., HAGEMAN R., REYNOLDS V., BOISVENUE R.J.,  
 RA COX G.N.;  
 RT "Cloning and sequence comparisons of four distinct cysteine proteases  
 RT expressed by Haemochus contortus adult worms.";  
 RL Mol. Biochem. Parasitol. 51:209-218(1992).  
 DR EMBL: M80388; AAA29178.1; -.  
 DR HSP; P07858; ICSB.  
 DR PFAM: PF00112; Peptidase\_C1; 1.  
 FT NON\_TER 1  
 FT CHAIN 84 341 CYSTEINE PROTEINASE.  
 SQ SEQUENCE 341 AA; 38342 MW; 2520FB2D CRC32;

Query Match 65.3%; Score 32; DB 5; Length 341;  
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLRLDER 10  
 |||||  
 DB 221 YRLRLDER 230

RESULT 12  
 Q02895 PRELIMINARY; PRT; 342 AA.  
 AC Q02895;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE LPG20P.  
 GN LPG20W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89345597.  
 RA GANDEL A., TZAGOLOFF A.;  
 RT "Homology of aspartyl- and lysyl-tRNA synthetases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6023-6027(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90078303.  
 RA DESHAIES R.J., SCHEKMAN R.;  
 RT "SEC62 encodes a putative membrane protein required for protein  
 RT translocation into the yeast endoplasmic reticulum.";  
 RL J. Cell Biol. 109:0-0(0).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94019235.  
 RA SHIRAYAMA M., KAWAKAMI K., MATSUI Y., TANAKA K., TOH-E A.;  
 RT "MS13, a multicopy suppressor of mutants hyperactivated in the RAS-  
 RT CAMP pathway, encodes a novel HSP70 protein of Saccharomyces  
 RT cerevisiae.";  
 RL Mol. Gen. Genet. 240:323-332(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96009602.  
 RA WATANABE Y., IRIE K., MATSUMOTO K.;  
 RT "Yeast RLMI encodes a serum response factor-like protein that may  
 RT function downstream of the Mpk1 (Slt2) mitogen-activated protein  
 RT kinase pathway.";  
 RL Mol. Cell. Biol. 15:5740-5749(1995).  
 RN [5]  
 RP SEQUENCE FROM N.A.

```

RX MEDLINE; 97313271.
RA BUSSY H., STORMS R.K., AHMED A., ALBERMANN K., ALLEN E., ANSORGE W.,
RA ARAUJO R., APARICIO A., BARRELL B., BADCOCK K., BENES V., BOTSTEIN D.,
RA BOWMAN S., BRUCKNER M., CARPENTER J., CHERRY J.M., CHUNG E.,
RA CHURCHER C., COSTER F., DAVIS K., DAVIS R.W., DIETRICH F.S.,
RA DELIUS H., DIPALO T., DUBOIS E., DUSTERHOFT A., DUNCAN M., FLOETH M.,
RA FORTIN N., FRIESEN J.D., FRITZ C., GOFFEAU A., HALL J., HEBLING U.,
RA HEUMANN K., HILBERT H., HILLIER L., HUNICKE-SMITH S., HYMAN R.,
RA JOHNSON M., KALMAN S., KLEINE K., KOMP C., KURDI O., LASHKARI D.,
RA LEW H., LIN A., LIN D., LOUIS E.J., MARATHE R., MESSENGUY F.,
RA MEWES H.W., MITIPATI S., MOESTIL D., MULLER-AUER S., NAMATH A.,
RA NENTWICH U., OEFNER P., PEARSON D., PETEL F.X., POHL T.M.,
RA PURNELLE D., SCHAFFER M., SCHERENS B., SCHRAMM S.,
RA SCHROEDER M., SDICU A.M., TETTEIN H., URRESTARAZU L.A., USHINSKY S.,
RA VIERENDELS F., VISSERS S., VOSS H., WALSH S.V., WAMBUTT R., WANG Y.,
RA WEDLER E., WEDLER H., WINNETT E., ZHONG W.W., ZOLLNER A., VO D.H.,
RA HANI J.,
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
RL Nature 387:0-0(0).
RN [6]
RN SEQUENCE FROM N.A.
RA AVRAM D.A., BAKALINSKY A.T.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RN SEQUENCE FROM N.A.
RA HILL J.E., TZAGOLOFF A.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [8]
RN SEQUENCE FROM N.A.
RA WANG Y., AHMED A., BUSSEY H., FORTIN N., FRIESEN J.D., HALL J.,
RA STORMS R.K., VO D.H., WINNETT E.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [9]
RN SEQUENCE FROM N.A.
RA COLLINSON L.P., DAWES I.W.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [10]
RN SEQUENCE FROM N.A.
RA BUSSEY H.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [11]
RN SEQUENCE FROM N.A.
RA JIA Y., CHERRY J.M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U43281; AAB68211.1; -.
SQ SEQUENCE 342 AA; 39682 MW; 9EAC1CD5 CRC32;

Query Match 65.3%; Score 32; DB 3; Length 342;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIIRLDER 10
DB 207 YNLLYREDER 216

RESULT 13
Q9X314 PRELIMINARY; PRT; 428 AA.
ID Q9X314
AC Q9X314;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PX01-44.
OS Bacillus anthracis.
OG Plasmid virulence plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-STERNE;
RA OKINAKA R.T., CLOUD K., HAMTON O., HOFFMASTER A., HILL K.K., KEIM P.,

```

```

RA KOEHLER T., LAMKE G., KOMANO S., MAHILLON J., MANTER D., MARTINEZ Y.,
RA RICKE D.O., SVENSSON R., JACKSON P.J.;
RT "The sequence and organization of pX01, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 0:0-0(1999).
DR EMBL; AF065404; AAD32348.1; -.
KW Plasmid.
SQ SEQUENCE 428 AA; 50415 MW; F0FE581D CRC32;

Query Match 65.3%; Score 32; DB 2; Length 428;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLIIRLDE 9
DB 347 YRMIALRDD 355

RESULT 14
O16954 PRELIMINARY; PRT; 441 AA.
ID O16954
AC O16954;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE SIMILARITY TO C4-TYPE ZINC FINGERS.
GN R1LG11.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718;
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SNALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA DAVIDSON S., WOHLDMANN P., BAUER C.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF022976; AAC69086.1; -.
DR PFAM; PF01362; DUF12.1; -.
SQ SEQUENCE 441 AA; 51624 MW; DE523A9B CRC32;

Query Match 65.3%; Score 32; DB 5; Length 441;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIIRLDER 10
DB 161 YRLTKIDQR 170

RESULT 15

```

O06421  
 ID O06421 PRELIMINARY; PRT; 554 AA.  
 AC O06421;  
 DT 01-JUL-1997 (TReMBLrel. 04, Created)  
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
 DE 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
 DE MEND.  
 GN MEND.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA BROWN D., CHURCHER C.M.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA BARRELL B.G., RAJANDREAM M.A.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE; 96181548.  
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RT "An integrated map of the genome of the tubercle bacillus,  
 RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium  
 RT leprae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).  
 DR EMBL: 295558; CAB08966.1; -;  
 SQ SEQUENCE 554 AA; 57835 MW; C42C89FC CRC32;

Query Match 65.3%; Score 32; DB 2; Length 554;  
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLLRLDER 10  
 || :|||  
 Db 48 RLHVRIDER 56

Search completed: February 8, 2000, 13:17:42  
 Job time: 32491 sec

OM of: US-08-653-294-15 to: GenEmbl:\* out\_format : pfs

Date: Feb 8, 2000 4:40 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frameat.p2n.model -DEV=xlp  
-O=Cgml\_1/USPTO.spool/US08653294/runat\_04022000\_160701\_15779/app\_query.fasta.1  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-15

Query length: 10

Database: GenEmbl:\*

Database sequences: 821193

Database length: -1518192014

Search time (sec): 11370.480000

score\_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_hg3:AC008537	-	42.00	100.64	1.4e-03	187246	..
gb_hg4:AC008591	-	41.00	95.48	2.7e-03	211769	! AC008591 Homo sapiens chromo
gb_pr2:HS1106N18	-	39.00	90.86	4.9e-03	142336	! AL033457 Human DNA sequence
gb_hg2:AC004965	+	39.00	83.57	1.2e-04	323732	! AC004965 Homo sapiens clone
gb_p11:SPBC14C8	+	38.00	99.49	1.6e-03	34051	! AL022305 S.pombe chromosome I
gb_in1:CELC24G7	-	38.00	96.63	2.3e-03	46996	! U88310 Caenorhabditis elegans
gb_hg5:AC014106	+	38.00	93.91	3.3e-03	63852	! AC014106 Drosophila melanog
gb_hg2:AC006902	+	38.00	90.25	5.3e-03	96468	! AC006902 Caenorhabditis eleg
gb_hg4:AC010050	-	38.00	88.48	6.5e-03	117732	! AC010050 Drosophila melanog
gb_hg4:AC008496	-	38.00	87.23	7.8e-03	135436	! AC008496 Homo sapiens chromo
gb_pr3:AC003082	+	38.00	85.21	1.0e-04	170136	! AC003082 Human BAC clone RGO
gb_hg2:AC006780	-	38.00	80.18	1.9e-04	299719	! AC006780 Caenorhabditis eleg
gb_hg2:AC006858	-	38.00	80.18	1.9e-04	299719	! AC006858 Caenorhabditis eleg
gb_p12:AF026087	-	37.00	119.52	1.23-53	2253	! AF026087 Schizosaccharomyces
gb_pr1:HUMM6PR	-	37.00	118.86	1.34-51	2428	! M16985 Human cation-dependent
gb_pr1:THMR465	-	37.00	116.98	1.71-44	3000	! X56257 Human MPR46 gene for 46
gb_ba1:THARGG	+	37.00	114.09	247.90	4154	! M32298 Thermoplasma acidophilu
gb_p11:AB011417	-	37.00	107.59	571.07	8646	! AB011417 Gibberella zeae genes
gb_ba2:AF061251	-	37.00	103.31	988.68	14002	! AF061251 Escherichia coli sero
gb_ba1:AB008676	-	37.00	95.11	2.5e-03	31482	! AB008676 Escherichia coli O15
gb_p12:SPBC23E5	-	37.00	95.31	2.8e-03	34486	! AL022387 S.pombe chromosome I
gb_hg3:AC008847	+	37.00	93.93	3.3e-03	40252	! AC008847 Homo sapiens chromos
gb_p11:AB008268	-	37.00	87.45	7.6e-03	83594	! AB008268 Arabidopsis thaliana
gb_hg5:AC0011643	-	37.00	82.19	1.5e-04	151177	! AC0011643 Homo sapiens clone
gb_pr4:AC006581	-	37.00	80.99	1.7e-04	172931	! AC006581 Homo sapiens 12p21
gb_ba1:AP000003	-	37.00	78.35	2.4e-04	233000	! AP000003 Pyrococcus horikosh
gb_in1:AF026152	+	36.00	129.07	36.32	486	! AF026152 Caenorhabditis elegans
gb_p11:CS15SRN1	-	36.00	117.90	152.13	1710	! Z35320 C.sericea (85.35) gene
gb_p11:SPU45981	-	36.00	117.00	170.79	1893	! U45981 Schizosaccharomyces pom
gb_om:RABUPNAA	-	36.00	114.58	232.81	2485	! M57301 O.cuniculus UDP-N-acety
gb_ba1:S68137	+	36.00	114.51	234.95	2505	! S68137 orf 5' of mpal, mpal-Pu
gb_ba1:PASMP	+	36.00	114.34	240.30	2555	! M93021 Pasteurella haemolytica
gb_pr2:HUMSAA	+	36.00	111.64	339.36	3460	! J03474 Human serum amyloid A g
gb_p11:ATHSP881	-	36.00	109.22	462.82	4544	! Y11829 A.thaliana hsp88.1 gene
gb_in1:CELC04F6	-	36.00	94.06	3.2e-03	25083	! U42835 Caenorhabditis elegans
gb_in1:CELC09F7	-	36.00	91.84	4.3e-03	32202	! U00050 Caenorhabditis elegans
gb_p12:SPAC1F5	-	36.00	89.86	5.5e-03	40242	! Z68136 S. pombe chromosome I
gb_hg4:AC010028	+	36.00	82.32	1.5e-04	94179	! AC010028 Drosophila melanog
gb_p12:AC0007167	+	36.00	81.45	1.6e-04	103874	! AC0007167 Arabidopsis thalian
gb_p12:ATAC006234	+	36.00	79.48	2.1e-04	129667	! AC006234 Arabidopsis thalian
gb_p11:AP000815	-	36.00	78.65	2.3e-04	142418	! AP000815 Oryza sativa genom
gb_hg6:AC012259	-	36.00	77.63	2.6e-04	159645	! AC012259 Homo sapiens chromo

gb\_hg4:AC009643 + 36.00 76.47 3.1e+04 182069 ! AC009643 Homo sapiens chr  
gb\_hg2:AC004618 + 36.00 75.63 3.4e+04 200000 ! AC004618 Homo sapiens chr  
gb\_hg2:AC004670 + 36.00 75.63 3.4e+04 200000 ! AC004670 Homo sapiens chr

seq\_name: gb\_hg3:AC008537

seq\_documentation\_block:

LOCUS AC008537 187246 bp DNA HTG 02-SEP-1999  
DEFINITION Homo sapiens chromosome 19 clone CIT-HSPC\_490E21, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 93 unordered pieces.

ACCESSION AC008537.1 GI:5686530

VERSION AC008537.1

KEYWORDS HTG; HTGS\_PHAISE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 187246)

Sequencing of Human Chromosome 19

Unpublished

2 (bases 1 to 187246)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

www.jgi.doe.gov.

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 93 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved. 651: contig of 651 bp in length

\* 1

\* 652 1286: contig of 635 bp in length

\* 1287 1986: contig of 700 bp in length

\* 1987 2249: contig of 263 bp in length

\* 2250 2892: contig of 643 bp in length

\* 2893 3502: contig of 610 bp in length

\* 3503 4010: contig of 508 bp in length

\* 4011 4658: contig of 648 bp in length

\* 4659 5315: contig of 657 bp in length

\* 5316 5944: contig of 629 bp in length

\* 5945 6344: contig of 400 bp in length

\* 6345 6384: contig of 40 bp in length

\* 6385 6641: contig of 257 bp in length

\* 6642 7288: contig of 647 bp in length

\* 7289 7963: contig of 675 bp in length

\* 7964 8224: contig of 261 bp in length

\* 8225 8497: contig of 273 bp in length

\* 8498 8570: contig of 73 bp in length

\* 8571 8645: contig of 75 bp in length

\* gap of unknown length

```
* 8646 9296: contig of 651 bp in length
* gap of unknown length
* 9297 9342: contig of 46 bp in length
* gap of unknown length
* 9343 9531: contig of 189 bp in length
* gap of unknown length
* 9532 10093: contig of 562 bp in length
* gap of unknown length
* 10094 10280: contig of 187 bp in length
* gap of unknown length
* 10281 10875: contig of 595 bp in length
* gap of unknown length
* 10876 11479: contig of 604 bp in length
* gap of unknown length
* 11480 12171: contig of 692 bp in length
* gap of unknown length
* 12172 12796: contig of 625 bp in length
* gap of unknown length
* 12797 12914: contig of 118 bp in length
* gap of unknown length
* 12915 12972: contig of 58 bp in length
* gap of unknown length
* 12973 13056: contig of 84 bp in length
* gap of unknown length
* 13057 13701: contig of 645 bp in length
* gap of unknown length
* 13702 14306: contig of 605 bp in length
* gap of unknown length
* 14307 14365: contig of 59 bp in length
* gap of unknown length
* 14366 14449: contig of 84 bp in length
* gap of unknown length
* 14450 14872: contig of 423 bp in length
* gap of unknown length
* 14873 14935: contig of 63 bp in length
* gap of unknown length
* 14936 14992: contig of 57 bp in length
* gap of unknown length
* 14993 15565: contig of 573 bp in length
* gap of unknown length
* 15566 16178: contig of 613 bp in length
* gap of unknown length
* 16179 16787: contig of 609 bp in length
* gap of unknown length
* 16788 17451: contig of 664 bp in length
* gap of unknown length
* 17452 18132: contig of 681 bp in length
* gap of unknown length
* 18133 18777: contig of 645 bp in length
* gap of unknown length
* 18778 19514: contig of 737 bp in length
* gap of unknown length
* 19515 20140: contig of 626 bp in length
* gap of unknown length
* 20141 20198: contig of 58 bp in length
* gap of unknown length
* 20199 20274: contig of 76 bp in length
* gap of unknown length
* 20275 20399: contig of 125 bp in length
* gap of unknown length
* 20400 20963: contig of 564 bp in length
* gap of unknown length
* 20964 21089: contig of 126 bp in length
* gap of unknown length
* 21090 21301: contig of 212 bp in length
* gap of unknown length
* 21302 21374: contig of 73 bp in length
* gap of unknown length
* 21375 21650: contig of 276 bp in length
* gap of unknown length
* 21651 22273: contig of 623 bp in length
* gap of unknown length
* 22274 22849: contig of 576 bp in length
```

```
* 22850 24258: contig of 1409 bp in length
* gap of unknown length
* 24259 25504: contig of 1246 bp in length
* gap of unknown length
* 25505 26750: contig of 1246 bp in length
* gap of unknown length
* 26751 27915: contig of 1165 bp in length
* gap of unknown length
* 27916 28174: contig of 259 bp in length
* gap of unknown length
* 28175 30152: contig of 1978 bp in length
* gap of unknown length
* 30153 31646: contig of 1494 bp in length
* gap of unknown length
* 31647 33042: contig of 1396 bp in length
* gap of unknown length
* 33043 34214: contig of 1172 bp in length
* gap of unknown length
* 34215 35095: contig of 881 bp in length
* gap of unknown length
* 35096 36881: contig of 1786 bp in length
* gap of unknown length
* 36882 39070: contig of 2189 bp in length
* gap of unknown length
* 39071 41053: contig of 1983 bp in length
* gap of unknown length
* 41054 43362: contig of 2309 bp in length
* gap of unknown length
* 43363 45689: contig of 2327 bp in length
* gap of unknown length
* 45690 47736: contig of 2047 bp in length
* gap of unknown length
* 47737 50962: contig of 3226 bp in length
* gap of unknown length
* 50963 54510: contig of 3548 bp in length
* gap of unknown length
* 54511 57987: contig of 3477 bp in length
* gap of unknown length
* 57988 60620: contig of 2633 bp in length
* gap of unknown length
* 60621 64843: contig of 4223 bp in length
* gap of unknown length
* 64844 68934: contig of 4091 bp in length
* gap of unknown length
* 68935 72268: contig of 3334 bp in length
* gap of unknown length
* 72269 76549: contig of 4281 bp in length
* gap of unknown length
* 76550 82005: contig of 5456 bp in length
* gap of unknown length
* 82006 86293: contig of 4288 bp in length
* gap of unknown length
* 86294 91795: contig of 5502 bp in length
* gap of unknown length
* 91796 98179: contig of 6384 bp in length
* gap of unknown length
* 98180 104674: contig of 6495 bp in length
* gap of unknown length
* 104675 114305: contig of 9631 bp in length
* gap of unknown length
```

## alignment\_scores:

Quality: 42.00 Length: 9  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-653-294-15 x AC008537/rev ..

Align seg 1/1 to reverse of: AC008537 from: 1 to: 187246

2 ArgLeuLeuIleArgLeuAspGluArg 10

```
|||||
4014 GCGCTTCATCAGGTGGACGAGA 3988

seq_name: gb_htg4:AC008591

seq_documentation_block:
LOCUS AC008591 211769 bp DNA HTG 31-OCT-1999
DEFINITION Homo sapiens chromosome 5 clone C17-HSPC_575N7, *** SEQUENCING IN
PROGRESS ***, 60 unordered pieces.
ACCESSION AC008591
VERSION AC008591.2 GI:6165161
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 211769)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 211769)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 31, 1999 this sequence version replaced gi:5686476.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 518: contig of 518 bp in length
* gap of unknown length
* 519 1379: contig of 861 bp in length
* gap of unknown length
* 1380 1992: contig of 613 bp in length
* gap of unknown length
* 1993 3019: contig of 1027 bp in length
* gap of unknown length
* 3020 4116: contig of 1097 bp in length
* gap of unknown length
* 4117 4468: contig of 352 bp in length
* gap of unknown length
* 4469 4889: contig of 401 bp in length
* gap of unknown length
* 4870 6986: contig of 2117 bp in length
* gap of unknown length
* 6987 8350: contig of 1364 bp in length
* gap of unknown length
* 8351 9169: contig of 819 bp in length
* gap of unknown length
* 9170 10576: contig of 1407 bp in length
* gap of unknown length
* 10577 10743: contig of 167 bp in length
* gap of unknown length
* 10744 11269: contig of 526 bp in length
* gap of unknown length
* 11270 11921: contig of 652 bp in length
* gap of unknown length
* 11922 12533: contig of 612 bp in length
* gap of unknown length
* 12534 12829: contig of 296 bp in length
* gap of unknown length
* 12830 13906: contig of 1077 bp in length
* gap of unknown length
* 13907 14105: contig of 199 bp in length
* gap of unknown length
* 14106 14914: contig of 809 bp in length
* gap of unknown length
* 14915 16489: contig of 1575 bp in length
* gap of unknown length

* 16490 16732: contig of 243 bp in length
* gap of unknown length
* 16733 16933: contig of 201 bp in length
* gap of unknown length
* 16934 17739: contig of 806 bp in length
* gap of unknown length
* 17740 17875: contig of 136 bp in length
* gap of unknown length
* 17876 19318: contig of 1443 bp in length
* gap of unknown length
* 19319 21247: contig of 1929 bp in length
* gap of unknown length
* 21248 22282: contig of 1035 bp in length
* gap of unknown length
* 22283 23232: contig of 950 bp in length
* gap of unknown length
* 23233 24221: contig of 989 bp in length
* gap of unknown length
* 24222 25716: contig of 1495 bp in length
* gap of unknown length
* 25717 27600: contig of 1884 bp in length
* gap of unknown length
* 27601 28647: contig of 1047 bp in length
* gap of unknown length
* 28648 30401: contig of 1754 bp in length
* gap of unknown length
* 30402 31662: contig of 1261 bp in length
* gap of unknown length
* 31663 34613: contig of 2951 bp in length
* gap of unknown length
* 34614 35537: contig of 924 bp in length
* gap of unknown length
* 35538 37384: contig of 1847 bp in length
* gap of unknown length
* 37385 38975: contig of 1591 bp in length
* gap of unknown length
* 38976 42433: contig of 3458 bp in length
* gap of unknown length
* 42434 46186: contig of 3753 bp in length
* gap of unknown length
* 46187 48955: contig of 2769 bp in length
* gap of unknown length
* 48956 52157: contig of 3202 bp in length
* gap of unknown length
* 52158 57534: contig of 5377 bp in length
* gap of unknown length
* 57535 60590: contig of 3056 bp in length
* gap of unknown length
* 60591 66089: contig of 5499 bp in length
* gap of unknown length
* 66090 70445: contig of 4356 bp in length
* gap of unknown length
* 70446 77265: contig of 6820 bp in length
* gap of unknown length
* 77266 82716: contig of 5451 bp in length
* gap of unknown length
* 82717 87967: contig of 5251 bp in length
* gap of unknown length
* 87968 95151: contig of 7184 bp in length
* gap of unknown length
* 95152 102796: contig of 7645 bp in length
* gap of unknown length
* 102797 111023: contig of 8227 bp in length
* gap of unknown length
* 111024 119350: contig of 8327 bp in length
* gap of unknown length
* 119351 129014: contig of 9664 bp in length
* gap of unknown length
* 129015 140858: contig of 11844 bp in length
* gap of unknown length
* 140859 154309: contig of 13451 bp in length
* gap of unknown length
```

The true left end of clone 908p16 is at 142337 in this sequence. The true right end of clone dJ119N1 is at 79367 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' character.

★ 7887 7905: g



\* 7906 9644: contig of 1739 bp in length  
\* 9645 gap of unknown length  
\* 9664 11563: contig of 1900 bp in length  
\* 11564 gap of unknown length  
\* 11583 13437: contig of 1855 bp in length  
\* 13438 13456: gap of unknown length  
\* 13457 15087: contig of 1631 bp in length  
\* 15088 15106: gap of unknown length  
\* 15107 17659: contig of 2553 bp in length  
\* 17660 17678: gap of unknown length  
\* 17679 19643: contig of 1965 bp in length  
\* 19644 19682: gap of unknown length  
\* 19683 21474: contig of 1812 bp in length  
\* 21475 21494: gap of unknown length  
\* 21494 23932: contig of 2499 bp in length  
\* 23933 24011: gap of unknown length  
\* 24012 27289: contig of 3258 bp in length  
\* 27290 27288: gap of unknown length  
\* 27289 30064: contig of 2776 bp in length  
\* 30065 30083: gap of unknown length  
\* 30084 33836: contig of 3753 bp in length  
\* 33837 33855: gap of unknown length  
\* 33856 36684: contig of 2829 bp in length  
\* 36685 36703: gap of unknown length  
\* 36704 39501: contig of 2798 bp in length  
\* 39502 39520: gap of unknown length  
\* 39521 43003: contig of 3483 bp in length  
\* 43004 43022: gap of unknown length  
\* 43023 46858: contig of 3836 bp in length  
\* 46859 46877: gap of unknown length  
\* 46878 50622: contig of 3745 bp in length  
\* 50623 50641: gap of unknown length  
\* 50642 54121: contig of 3480 bp in length  
\* 54122 54140: gap of unknown length  
\* 54141 58079: contig of 3939 bp in length  
\* 58080 58098: gap of unknown length  
\* 58099 61695: contig of 3597 bp in length  
\* 61696 61714: gap of unknown length  
\* 61715 66025: contig of 4311 bp in length  
\* 66026 66044: gap of unknown length  
\* 66045 67681: contig of 1637 bp in length  
\* 67682 67700: gap of unknown length  
\* 67701 70508: contig of 2808 bp in length  
\* 70509 75295: contig of 4768 bp in length  
\* 75296 75314: gap of unknown length  
\* 75315 80384: contig of 5050 bp in length  
\* 80385 80383: gap of unknown length  
\* 80384 82047: contig of 1664 bp in length  
\* 82048 82066: gap of unknown length  
\* 82067 88837: contig of 6761 bp in length  
\* 88838 88846: gap of unknown length  
\* 88847 98782: contig of 9916 bp in length  
\* 98783 98781: gap of unknown length  
\* 98782 110167: contig of 11386 bp in length  
\* 110168 110186: gap of unknown length  
\* 110187 122949: contig of 12763 bp in length  
\* 122950 122968: gap of unknown length  
\* 122969 140287: contig of 17299 bp in length  
\* 140288 140286: gap of unknown length  
\* 140289 163837: contig of 23551 bp in length  
\* 163838 163856: gap of unknown length  
\* 163857 192172: contig of 28316 bp in length  
\* 192173 192191: gap of unknown length  
\* 192192 228203: contig of 36012 bp in length  
\* 228204 228222: gap of unknown length  
\* 228223 317448: contig of 89226 bp in length  
\* 317449 317466: gap of unknown length  
\* 317467 318949: contig of 1483 bp in length  
\* 318950 318967: gap of unknown length  
\* 318968 320503: contig of 1536 bp in length  
\* 320504 320521: gap of unknown length  
\* 320522 322060: contig of 1539 bp in length

\* 322061 322078: gap of unknown length  
\* 322079 323792: contig of 1714 bp in length.  
FEATURES  
source  
1..323792  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DJ1106H14"  
BASE COUNT 81610 a 77790 c 79384 g 84233 t 775 others  
ORIGIN  
alignment\_scores:  
Quality: 39.00 Length: 10  
Ratio: 4.333 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 80.000  
alignment\_block:  
US-08-653-294-15 x AC004965 ..  
Align seg 1/1 to: AC004965 from: 1 to: 323792  
1 TyrArgLeuLeuLeuArgLeuAspGluArg 10  
||||||| ||||||| ||||||| ||||||| |||||||  
75519 TATAGACTGTATATCAGGCTGGACACAGG 75548  
seq\_name: gb\_pl1:SPBC14C8  
seq\_documentation\_block:  
LOCUS SPBC14C8 34051 bp DNA PLN 19-OCT-1998  
DEFINITION S.pombe chromosome II cosmid cl4C8.  
ACCESSION AL022305  
VERSION AL022305.1 GI:3647356  
KEYWORDS acetolactate synthase regulatory subunit; beta transducin; cdc18;  
cell division control protein; cut2; DNA polymerase V; G-beta  
repeat; glucosyltransferase precursor; methionine aminopeptidase; pol5;  
protein import protein; ribosomal protein; RNA polymerase II  
subunit; rpb8; transcription factor; triacylglycerol lipase; WD  
domain.  
SOURCE fission yeast.  
ORGANISM Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomyces.  
REFERENCE 1 (bases 1 to 34051)  
AUTHORS Lyne,M., Rajandream,M.A., Barrell,B.G. and Volckaert,G.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAR-1998) European Schizosaccharomyces genome  
sequencing project, Sanger Centre, The Wellcome Trust Genome  
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barreillesanger.ac.uk  
and Katholieke Universiteit Leuven, Faculty of Agricultural and  
Applied Biological Sciences, Laboratory of Gene Technology,  
Kardinaal Mercierlaan 92 Blok F, B-3001 Leuven, Belgium  
REFERENCE 2 (bases 18652 to 34051)  
Xiang,Z. and Aves,S.  
AUTHORS Unpublished  
JOURNAL  
REMARK Department of Biological Sciences, University of Exeter, Perry  
Road, Exeter EX4 4QG, United Kingdom  
COMMENT On Sep 25, 1998 this sequence version replaced gi:3006158.  
Notes:  
Protein coding regions (CDS) have been predicted with the help of  
computer analysis using the Genefinder program in PomBase (an ACEDB  
database) with additional predictions for the branch-acceptor sites  
supplied by the program Splice. CAUTION: It is possible that for  
any individual CDS we may have underestimated or overestimated the  
number of introns/exons or we may not have chosen the correct  
splice donor/acceptor sites. CDS are numbered using the following  
system eg SPBC25H2.01c.sp (S. pombe), B (chromosome 2), c25H2  
(Cosmid name), .01 (first CDS), C (complementary strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid cl4C8 is overlapped at the 3' end by cl5C4. Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.

(URL, [http://www.sanger.ac.uk/Projects/S\\_pombe/](http://www.sanger.ac.uk/Projects/S_pombe/))  
During 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.

## FEATURES

## source

1. 34051

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db\_xref="taxon:4896"

/chromosome="II"

/clone="cosmid cl4C8"

/map="IIR"

/complement(1..661)

/gene="cut2"

/complement(<1..661)

/gene="cut2"

/note="SPBC14C8.01c, len:>220"

/codon\_start=1

/product="spindle pole body regulator"

/protein\_id="CAA18419.1"

/db\_xref="GI:3006159"

/translation="MLPRTMFSGYKNAFVPTIPNNGTKGACSKRAPLGSTQKSA

PSSVTPRTVJGGSTNISKFSIPASTKMSPMDSIPSTPILEPNISOGISRSVQR  
SKRLASPRSLSDTLPPLNELEDIEYPPPHLDPIQSLGFDVDAICETLDPWS  
MKNKATSVIRNTPASDFHYKEFSDDDPIDQPLLSVDGDSPLTEKDTNLTTPATLKA  
SD"

1994. 3277

/gene="SPBC14C8.02"

1994..3277

/gene="SPBC14C8.02"

/note="SPBC14C8.02, len:427, SIMILARITY: Saccharomyces

cerevisiae, IM44.YEAST, mitochondrial import inner

membrane translocase subunit tim44 precursor, (431 aa),

fasta scores: opt: 893, E():0, (43.0% identity in 430 aa)"

/codon\_start=1

/product="mitochondrial import inner membrane translocase

subunit precursor"

/protein\_id="CAA18420.1"

/db\_xref="GI:3006160"

/db\_xref="SPTREMBL:O60084"

/translation="MFLSRKITKARTIVQCFQFQIFSQNAPOSPKVFMDTFR

AELKSOELQSVLQDSSGSLSESDTFKKARDAYEKARSGTTAASSFGTKVGRAG

AKIGSYAKAWESAPVQLSKVISSANTATVATGYDTATKPVRETAFYTKIKOTMSDGS

TSSRYGYADKEQKRLKEEPERRNMFASARIQPNEDVQSVVHVHNPNSWKNVEQI

KNESRLVKIQIELKSYOESHPVSVIRDMADISIGVWSRMFSEASQVMRRFKEI

DPSTNTEHFLQYREYIVPEVTEAYVKGDKVELATLWSEAPFSVYETITKEYAKHGV

SVGKILDRGVDSIQRLQPNIDPVIFIVTRQEVHMFQDASSGELVAGKDDRIQOC

TYASVFRVEDELDPNTRGRVIYDFARAVADF"

2261..2284

## misc\_feature

/gene="SPBC14C8.02"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

4172..5452

/gene="SPBC14C8.03"

4172..5452

/gene="SPBC14C8.03"

/note="SPBC14C8.03, len:426, SIMILARITY: Arabidopsis

thaliana, G3212871, putative translation initiation

factor., (431 aa), fasta scores: opt: 1697, E():0, (59.4%

identity in 416 aa)"

/codon\_start=1

/product="putative methionine metalloproteinase"

/protein\_id="CAA18421.1"

/db\_xref="GI:3006161"

/db\_xref="SPTREMBL:O60085"

/translation="WTSATTEATPAKLOELKSLKENDVVEDDGKVEENDAAREGASN

GEKKKKKKKKTKTPQEQTNPTVGLSKIFVNNKKYPVGEVCDAEDNLWRTTDEE

KRALDRONFDQNDLRRAAEVHQARQYAGSVIKPGMSMDVWNTIENTTRALVEEDG

LKSGIGPTGVSILNHCAAHYTPNAGDTTILKEKDVVKVMDIGVHVNGRIVDSAFSTMSFD  
PDYNLAAVKAATNGIEEAGIDARLAEIGALQEMYESYEVEINGKTHQVKSIRNL  
CGNLDYIIHGKSGVPIVKGGEIKMEGEIFAIEFTFGSTGRGVVHDEMECSHYAKI  
PDAGHIPRLPRAKALLTITQNGFTLPFCRRYLDRIGESKYLALNNLVSAAGIVQDY  
PLPDCIRGYSYTAQFEHTIILHPTQKEVVSRRDDY"

## misc\_feature

/gene="SPBC14C8.03"

/note="Pfam match to entry PF00557 pep\_M24,

metalloproteinase family M24"

4751..4801

/gene="SPBC14C8.03"

/note="PS01202 Methionine aminopeptidase subfamily 2

signature"

7258..8036

/gene="SPBC14C8.04"

join(7258..7445,7526..8036)

/gene="SPBC14C8.04"

/note="SPBC14C8.04, len:232, SIMILARITY: Saccharomyces

cerevisiae, YCL009c, ILV6.YEAST, acetolactate synthase

small subunit precursor, (309 aa), fasta scores: opt: 513,

E():0, (49.4% identity in 243 aa)"

/codon\_start=1

/product="acetolactate synthase small subunit precursor"

/protein\_id="CAA18422.1"

/db\_xref="GI:3006162"

/db\_xref="SPTREMBL:O60086"

/translation="MSTFNRVQRPRKRVNCLVQNEPGVLSGLIARGNIDSL

VVCATEVNSRMTIVLRGADEVQAKROIEDIVSVMAVDITGTSMVRELLAKV

SLIGDPHFHFSEKVAESTNAKAGDEGVNNAALQRLASQALAIQLTTLFH

GRVADISTETIILELTATPDVDNFLSLRPYGLACRTGTSAITRAPHNEVTEEA

EDDVEVEVLPFG"

7446..7451

/gene="SPBC14C8.04"

/note="gtatgt, splice donor sequence"

7511..7525

/gene="SPBC14C8.04"

/note="ttaacatttttag, splice branch and acceptor"

complement(8460..9812)

/gene="SPBC14C8.05c"

complement(8460..9812)

/gene="SPBC14C8.05c"

/note="SPBC14C8.05c, len:450, SIMILARITY: Aspergillus

shirousami, AMYG\_ASPSH, glucosylase precursor, (639 aa),

fasta scores: opt: 797, E():0, (39.5% identity in 446 aa)"

/codon\_start=1

/product="glucosylase precursor"

/protein\_id="CAA18423.1"

/db\_xref="GI:3006163"

/db\_xref="SPTREMBL:O60087"

/translation="MRTYVLFLLGVSVAESLLSPNKRKSEASMDWTDOQKGIAMG

HMLNIGDSMHAKDINPGCIASPSDSDPYVQWRDSALTIMTILDRFFSGDKGL

EPILVYKDMEMVRLQVNPSPGDFAGGLGEPKFNVDGTSYDGDWGRPNQDSFALRAI

AFIKYNYLFENGEVHEVTWIEAVLADLTANTHWETASPDMEIKDKVHTFTLAV

QKRAMQDGTAFAKRIGAPQAALYQRTIETIDILKGEFDPGNGVIGKGYDRSGSL

DCSTLLASLYNEFDMHILTLKQETRTDTPVYVQWQKQMGRTPEYDGVGSKSI

GNPWFCTSSAAEIIYKATAYDNKGLPELTENYHFNKFAEFGPGYNNVTRKKNH

"

TYADNLFKVAEEHPHNGSMSEQFSRDDGHQKGRDLTWSYSSLLNAYIRREAIKGSV

complement(8604..9725)

/gene="SPBC14C8.05c"

/note="Pfam match to entry PF00723 glycosyl\_hydrol,

Glycosyl hydrolases family 15"

10878..12058

/gene="sop2"

join(10878..10950,10998..12058)

/gene="sop2"

/note="SPBC14C8.06, len:377"

/codon\_start=1

/product="probable arp2-3 complex 41 kd subunit"

/protein\_id="CAA18424.1"

/db\_xref="GI:3006164"

/translation="MATSQVLHLPKPSYEHAFNSQRTSEFTVTTTATNQVELYEQDNGG



```

gene
28911..32411
/genes="C24G7.4"
join(28911..28981,29343..29464,29894..30247,30296..30386,
30573..30894,30943..31108,31289..31944,32202..32411)
/genes="C24G7.4"
/notes="similar to degenerins; coded for by C. elegans cDNA
C5MSH05F; coded for by C. elegans cDNA yk35e3.3; coded for
by C. elegans cDNA yk35e3.5"
/codon_start=1
/protein_id="AAB42340.1"
/db_xref="GI:1825743"
/translation="MFQKSYFNLSLSSLSAIFKAAKQFANQVLIQVFNVSF
KGLKIKGVSAFRKTHESSTTHGPKRIYFGKGVARAFWMLIVGLALMLCFQIF
ILQWYFSKPTLSQVSVFVNEGMDPPAVTVNENPIKSYVRELNVSGDLNGELFEY
LLQTNDAWFLSNLDRHLNKHETHDEATYFQNHDFDQIIKFLRAGDCGIFPCCCK
YKQKYSIGKWEELDLNLAPEWRRKQISPSSEAGLQIVDAQEEELKGNENDAKA
IFSDIYENGFRIYHPPTNAQLSEGISVSPSRVYSIAKTVTHNLNRGNWNCSE
NPEGYNTFLSYASACRALCIAQFNDTCGAPFTYVDRKKICAPYESTCDNH
MLKVNQDYLLEPDCIECHMECQSTYSYNSYGDGFGNRSGLWKLITSNKSETHIK
NVAVINIFLEMYTSYSQVQATSLTEILSDIGNMGMFLGMSVITITELSFTHIK
FWMVSKRQRYMSKTHKEKEHQLDEAVKEFOERSRNRSENISALGHYSRIT
PVDDQTKCYKNASEGMSNSSLDSVWELKEDINELRQLNQFSTQDIARILRPHQ
TSRONSTENYSSPSPIFTIEPMRSKQSKTSLPSSLSR"
32635..32636
/notes="see EST yk35e3.5"
32868..33569
/genes="C24G7.3"
join(32868..32898,32979..33141,33227..33569)
/genes="C24G7.3"
/notes="similar to mariner transposase"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42339.1"
/db_xref="GI:1825745"
/translation="MKLLGWRKILPDGWLGRLLIENKRVLYANHHTQAQWIGSGQT
ADQVPLKLRSNKVSFLGRETFDRSVLQDKKVFQHNNAQPYVTKHYKEDLAKCGW
PILTOPPYSPDISPSPDWLLNLTCALGRSPFTRENLTQDIELYFESLSAGFYRQGI
HKLNECWQDIVEDHGTYN"
38014..42000
/genes="C24G7.2"
join(38014..38120,38786..38931,38976..39021,39089..39197,
39250..39337,39384..39615,40120..40438,40757..40922,
40978..41046,41091..41253,41342..41433,41480..41625,
41677..41744,41863..42000)
/genes="C24G7.2"
/notes="similar to degenerins"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42338.1"
/db_xref="GI:1825744"
/translation="MEPTLSPNRYNEAFPHDDSYLVNFVAGSSGSESTPPSPFVPKC
NFRYNRSQMIIEVPVQAKLRLKLGTVSIKRETHCEFTTHGPKRIYFGKQKWA
TIFWLLVSCSLGLLITQVFIASELTSKPTVSDVSVFLNEDMDPFLITICNLAPIR
KYVNEINKTQSPVPMINMYMKWFEITPLIGGADRPTEHGEDELKLYMKHNLFT
VDSFMNSGFCSPDIFKLCGFCQEIFDCCTLSTEVLPKGCFTLDSSTKRAMHQ
TEPGIOAGLAILDLAHEQFDGSDMALFTNSFVNGFRVYFHPNPIPHLSDEFT
VSPNTVAFSAISSDRVLLPHTOWGNCNTEFPDQISLSSYSSGNCLSCKAKFYMEN
CGCTALYNIENNEKCTPYETTCLDNLAKPKNETGKIERQTPNCKACQACQNSLV
YRAYSIGSQFAGAFHYLKSINPEWDGMRANFQMINIFRDMSYEYNGVQDASV
TQLSDIGNMGMFLGMSVITITELCFKMPFLGFSKKRRLYFIRDMSYEYNGVQDASV
VCETVEKMAIASQGNLSLILKINKNYSEILNGSLVILSSRFTYPEHYFIENE
NILDF"
43708..46686
/genes="C24G7.1"
join(43708..43781,43825..43958,44004..44049,44094..44202,
44312..44399,44447..44678,44726..45044,45323..45488,
45545..45610,45653..45815,46110..46201,46250..46395,
46445..46512,46563..46686)
/genes="C24G7.1"
/notes="similar to degenerins"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42337.1"
/translation="MIPITISKPTNOSTKARKSNESYPSVTPRPNSSLHLIEV
PVALKKPKHAGETVSVVEREQHCETTHMGPKRIFOGKRWALFLWIMVCTSIGLL
PVOICILASNLSLPVSDVSEFLNEEGIQPQITICNFTPIRKTFYENMKTKGQISP
NNIMYIMHFTVPIPLIGSNWQLHEGKDLQEQYKNNPMTVOGFFIDAGFSCSDI
FKLCSFOGETDCGSIPTVPLGKCVTLDDLSTKPSMHKOTEPGIOAGLAILDA
HLEQFDGSDMALFTNSFVNGFRVYFHPNPIPHLSDEFTVTPNSVAYTAISSER
FELLTPNKGWNCTEHYPSGIKSDLPYLTGNCVSLCKAFEMENCGCTPAVYNERNLK

```

alignment\_scores:  
Quality: 38.00 Length: 10  
Ratio: 3.800 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 70.000

alignment\_block:  
US-08-653-294-15 x CELC24G7/rev ..  
Align seg 1/1 to reverse of: CELC24G7 from: 1 to: 46996  
1 TyxArgLeuLeuLeuAArgLeuAspGluArg 10  
|||||  
31227 TACAGTTATGATTCGTATGAGACACAGA 31298  
seq\_name: gb\_htg5:AC014106

seq\_documentation\_block:  
LOCUS AC014106 63852 bp DNA HTG 16-NOV-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces.  
ACCESSION AC014106  
VERSION AC014106.1 GI:6437229  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 63852)  
AUTHORS Adams,M. and Venter,J.C.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT This sequence was identified as CDM:10211010 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
Location/Qualifiers  
1..63852  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 18466 a 12714 c 12710 g 19962 t  
ORIGIN

alignment\_scores:  
Quality: 38.00 Length: 10  
Ratio: 3.800 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 70.000

alignment\_block:  
US-08-653-294-15 x AC014106 ..  
Align seg 1/1 to: AC014106 from: 1 to: 63852  
1 TyxArgLeuLeuLeuAArgLeuAspGluArg 10  
|||||  
36495 TATCGACTTTTAAATGCAATTAGATAAGAGA 36524  
seq\_name: gb\_htg2:AC006902

seq\_documentation\_block:

```

seq_name: gb_htg4:AC010050
seq_documentation_block:
LOCUS AC010050 117732 bp DNA HTG 16-OCT-1999
DEFINITION Drosophila melanogaster chromosome 3L/17L1 clone RPC198-21K15, ***
SEQUENCING IN PROGRESS ***, 23 unordered pieces.
ACCESSION AC010050.3 GI:6056153
VERSION AC010050
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 117732)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J.J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forecum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Osval,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J.J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Taber,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 117732)
Worley,K.C.
Direct Submission
Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 16, 1999 this sequence version replaced gi:5881454.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 838: contig of 838 bp in length
* 839 1645: contig of 807 bp in length
* 1646 2492: contig of 847 bp in length
* 2493 3485: contig of 993 bp in length
* 3486 4351: contig of 866 bp in length
* 4352 6048: contig of 1697 bp in length
* 6049 7829: contig of 1781 bp in length
* 7830 9752: contig of 1923 bp in length
* 9753 12320: contig of 2568 bp in length
* 12321 14723: contig of 2403 bp in length
* 14724 17531: contig of 2808 bp in length
* 17532 20954: contig of 3423 bp in length
* 20955 24775: contig of 3821 bp in length
* 24776 29461: contig of 4685 bp in length
* 29461 32706: contig of 3246 bp in length
* 32707 38082: contig of 5376 bp in length
* 38083 46290: contig of 8208 bp in length
* 46291 54186: contig of 7896 bp in length
* 54187 65008: contig of 11822 bp in length
* 65009 78413: contig of 13405 bp in length
* 78414 90799: contig of 11385 bp in length
* 90799 101691: contig of 10893 bp in length
* 101692 117732: contig of 16041 bp in length
*
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

seq_name: gb_htg4:AC010050
seq_documentation_block:
LOCUS AC010050 117732 bp DNA HTG 16-OCT-1999
DEFINITION Drosophila melanogaster chromosome 3L/17L1 clone RPC198-21K15, ***
SEQUENCING IN PROGRESS ***, 23 unordered pieces.
ACCESSION AC010050.3 GI:6056153
VERSION AC010050
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 117732)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J.J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forecum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Osval,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J.J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Taber,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 117732)
Worley,K.C.
Direct Submission
Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 16, 1999 this sequence version replaced gi:5881454.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 838: contig of 838 bp in length
* 839 1645: contig of 807 bp in length
* 1646 2492: contig of 847 bp in length
* 2493 3485: contig of 993 bp in length
* 3486 4351: contig of 866 bp in length
* 4352 6048: contig of 1697 bp in length
* 6049 7829: contig of 1781 bp in length
* 7830 9752: contig of 1923 bp in length
* 9753 12320: contig of 2568 bp in length
* 12321 14723: contig of 2403 bp in length
* 14724 17531: contig of 2808 bp in length
* 17532 20954: contig of 3423 bp in length
* 20955 24775: contig of 3821 bp in length
* 24776 29461: contig of 4685 bp in length
* 29461 32706: contig of 3246 bp in length
* 32707 38082: contig of 5376 bp in length
* 38083 46290: contig of 8208 bp in length
* 46291 54186: contig of 7896 bp in length
* 54187 65008: contig of 11822 bp in length
* 65009 79413: contig of 13405 bp in length
* 79414 90799: contig of 11385 bp in length
* 90799 101691: contig of 10893 bp in length
* 101692 117732: contig of 16041 bp in length
* 101693
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source
1..96468
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y74Ally"
BASE COUNT 30222 a 17588 c 18100 g 30408 t 150 others
ORIGIN
alignment_scores:
Quality: 38.00 Length: 10
Ratio: 3.800 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 70.000
alignment_block:
US-08-653-294-15 x AC006902/rev ..
Align seg 1/1 to reverse of: AC006902 from: 1 to: 96468
1 TyrArgLeuLeuLeuArgLeuAspGluArg 10
|||||
44659 TACAGGTATTGATTCGTATGGACACAGA 44630

```

```

FEATURES
  source
    1. 117732 Location/Qualifiers
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /chromosome="3L/71F1"
      /clone="RPC198-21K15"
BASE COUNT 36221 a 22525 c 23405 g 35555 t 26 others
ORIGIN

alignment_scores
  Quality: 38.00 Length: 10
  Ratio: 3.800 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
  US-08-653-294-15 x AC010050/rev ..
  Align seg 1/1 to reverse of: AC010050 from: 1 to: 117732
    1 TyrArgLeuLeuIleArgLeuAspGluArg 10
    |||||
    99919 TATCGACTTTTAAATGCAATAGATAAGAGA 99890

seq_name: gb_htg4:AC008496

seq_documentation_block:
  LOCUS AC008496 135436 bp DNA HTG 31-OCT-1999
  DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC_431G16, *** SEQUENCING IN
  PROGRESS ***, 6 ordered pieces.
  ACCESSION AC008496
  VERSION AC008496.2 GI:6165177
  KEYWORDS HTG: HTGS-PHASE2.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 135436)
    AUTHORS DOE Joint Genome Institute.
    TITLE Sequencing of Human Chromosome 5
    JOURNAL Unpublished
  REFERENCE
    2 (bases 1 to 135436)
    AUTHORS DOE Joint Genome Institute.
    TITLE Direct Submission
    JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
    Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  COMMENT
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 6 contigs. Gaps between the contigs
    * are represented as runs of N. The order of the pieces
    * is believed to be correct as given, however the sizes
    * of the gaps between them are based on estimates that have
    * provided by the submitter.
    * This sequence will be replaced
    * by the finished sequence as soon as it is available and
    * the accession number will be preserved
    * 1 32143: contig of 32143 bp in length
    * 32144 42505: gap of unknown length
    * 42506 76913: contig of 34408 bp in length
    * 76914 94682: gap of unknown length
    * 94683 122423: contig of 17769 bp in length
    * 122424 135436: gap of unknown length
    * Location/Qualifiers
    1. 135436
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CIT-HSPC_431G16"

FEATURES
  source
    1. 135436 Location/Qualifiers
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CIT-HSPC_431G16"

```

---

```

BASE COUNT 40855 a 29385 c 28075 g 36803 t 318 others
ORIGIN

alignment_scores
  Quality: 38.00 Length: 10
  Ratio: 3.800 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
  US-08-653-294-15 x AC008496/rev ..
  Align seg 1/1 to reverse of: AC008496 from: 1 to: 135436
    1 TyrArgLeuLeuIleArgLeuAspGluArg 10
    |||||
    10950 TTTCACCTACTCTTAGCTAGATGAGAGG 10921

seq_name: gb_pr3:AC003082

seq_documentation_block:
  LOCUS AC003082 170136 bp DNA PRI 06-NOV-1997
  DEFINITION Human BAC clone RG011J21 from 7q31, complete sequence.
  ACCESSION AC003082
  VERSION AC003082.1 GI:2588625
  KEYWORDS HTG.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 170136)
    AUTHORS Graves, T., Becker, M. and Hawkins, M.
    TITLE The sequence of H. sapiens BAC clone RG011J21
    JOURNAL Unpublished (1997)
  REFERENCE
    2 (bases 1 to 170136)
    AUTHORS Waterston, R.
    TITLE Direct Submission
    JOURNAL Submitted (06-NOV-1997) Department of Genetics, Washington
    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
  COMMENT
    SUBMITTED BY:
    Genome Sequencing Center
    Department of Genetics
    Washington University
    St. Louis MO 63108, USA
    http://genome.wustl.edu/gsc
    mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/dnr/gtb/chr7 or send
mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
Clone RG011J21 is from the first release of the human BAC library
CITB-978SK-B. The library contains cloned DNA from the male
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.

```

Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
 VECTOR: pBelOBAC11  
 Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RG067E13, 200 bp overlap.  
 Actual start of this clone is at base position 196 of RG011J21;  
 actual end is at 170136 of RG011J21.

This clone contains STS SWSS2102 (NID:g940333).

#### FEATURES

source	Location/Qualifiers	
	1..170136	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/clone="RG011J21"	
	/map="7q31"	
	complement(1..93)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	85..2579	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(1164..1582)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(2451..2509)	
misc_feature	/note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"	
misc_feature	complement(2451..2564)	
misc_feature	/note="similar to EST AA489997 (NID:g2218599) aa54c11.s1"	
misc_feature	2451..2528	
repeat_region	/note="similar to EST AA366493 (NID:g2018832)"	
repeat_region	complement(4227..4405)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(4719..4845)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(4929..5336)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(5525..5543)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	8635..8923	
repeat_region	/rpt_family="ALU"	
repeat_region	8924..8948	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(15996..16123)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	16125..21465	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(21024..21439)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	22327..23337	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(23173..23302)	
repeat_region	/note="similar to EST AA489997 (NID:g2218599) aa54c11.s1"	
repeat_region	complement(23173..23244)	
repeat_region	/note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"	
repeat_region	23184..23261	
repeat_region	/note="similar to EST AA366493 (NID:g2018832)"	
repeat_region	complement(27159..27188)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	28018..28045	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(28082..28155)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	28086..28175	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	28200..28284	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(29362..29403)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	29513..29694	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	29943..30142	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	30183..30377	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	3080..31186	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	31384..31422	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	31441..31480	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	33572..33606	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	34091..34382	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(34850..35137)	
repeat_region	/rpt_family="ALU"	
repeat_region	35855..36007	
repeat_region	/note="similar to EST AA425675 (NID:g2106395) zw47b05.s1"	
repeat_region	complement(38607..39020)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	39165..39186	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(41821..41959)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	41957..43047	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	43072..43358	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	43364..43656	
repeat_region	/rpt_family="ALU"	
repeat_region	43682..44482	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(44490..44609)	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	44565..45946	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(48709..48999)	
repeat_region	/rpt_family="ALU"	
repeat_region	49881..49901	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	51568..56075	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(52387..52447)	
repeat_region	/note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"	
repeat_region	52387..52466	
repeat_region	/note="similar to EST AA366493 (NID:g2018832)"	
repeat_region	complement(52387..52480)	
repeat_region	/note="similar to EST AA489997 (NID:g2218599) aa54c11.s1"	
repeat_region	56076..56419	
repeat_region	/rpt_family="MER"	
repeat_region	56420..56942	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(57178..57226)	
repeat_region	/rpt_family="ALU"	
repeat_region	59193..59213	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	complement(60100..61153)	
repeat_region	/rpt_family="MER"	
repeat_region	63986..64015	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	64113..64253	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(65648..65939)	
repeat_region	/rpt_family="ALU"	
repeat_region	66049..66338	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	66424..66798	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	66819..66995	
repeat_region	/rpt_family="L1"	/rpt_family="L1"
repeat_region	67032..67335	
repeat_region	/rpt_family="ALU"	
repeat_region	67336..68129	
repeat_region	/rpt_family="L1"	/rpt_family="L1"

```

repeat_region 69165..69708
               /rpt_family="MER"

alignment_scores:
  Quality: 38.00      Length: 10
  Ratio: 3.800       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 60.000

alignment_block:
US-08-653-294-15 x AC003082 ..
Align seg 1/1 to: AC003082 from: 1 to: 170136

1 TyArgLeuLeuIleArgLeuAspGluArg 10
|||||:|||||:|||||:|||||:|||||:
41782 TACAAACTTTAGTTAGTACGAGGAAAA 41811

seq_name: gb_htg2:AC006780

seq_documentation_block:
LOCUS AC006780 299719 bp DNA HTG 25-FEB-1999
DEFINITION Caenorhabditis elegans clone Y47D9A, *** SEQUENCING IN PROGRESS
ACCESSION AC006780
VERSION AC006780.2 GI:4309792
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
Waterston, R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 299719)
AUTHORS Waterston, R.H.
JOURNAL Direct Submission
TITLE Submitted (23-FEB-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 3047: contig of 3047 bp in length
* 3048 3061: gap of unknown length
* 3062 6816: contig of 3755 bp in length
* 6817 6830: gap of unknown length
* 6831 35765: contig of 28935 bp in length
* 35766 35779: gap of unknown length
* 35780 299719: contig of 263940 bp in length.
FEATURES
Source
1..299719
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y47D9A"

BASE COUNT 95763 a 55063 c 53844 g 95007 t 42 others
ORIGIN

alignment_scores:
  Quality: 38.00      Length: 10
  Ratio: 3.800       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 70.000

alignment_block:
US-08-653-294-15 x AC006780/rev ..
Align seg 1/1 to reverse of: AC006780 from: 1 to: 299719

1 TyArgLeuLeuIleArgLeuAspGluArg 10
|||||:|||||:|||||:|||||:|||||:
163290 TACAGGTATTGATTCGTATGGAGACAGA 163261

seq_name: gb_p12:AF026087

seq_documentation_block:
LOCUS AF026087 2253 bp DNA PLN 13-NOV-1997
DEFINITION Schizosaccharomyces pombe Sati gene, partial cds.
ACCESSION AF026087

```

```

Align seg 1/1 to reverse of: AC006780 from: 1 to: 299719

1 TyArgLeuLeuIleArgLeuAspGluArg 10
|||||:|||||:|||||:|||||:|||||:
163290 TACAGGTATTGATTCGTATGGAGACAGA 163261

seq_name: gb_htg2:AC006858

seq_documentation_block:
LOCUS AC006858 299719 bp DNA HTG 25-FEB-1999
DEFINITION Caenorhabditis elegans clone Y47D9X, *** SEQUENCING IN PROGRESS
ACCESSION AC006858
VERSION AC006858.1 GI:4263490
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
Waterston, R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 299719)
AUTHORS Waterston, R.H.
JOURNAL Direct Submission
TITLE Submitted (24-FEB-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 3047: contig of 3047 bp in length
* 3048 3061: gap of unknown length
* 3062 6816: contig of 3755 bp in length
* 6817 6830: gap of unknown length
* 6831 35765: contig of 28935 bp in length
* 35766 35779: gap of unknown length
* 35780 299719: contig of 263940 bp in length.
FEATURES
Source
1..299719
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y47D9X"

BASE COUNT 95763 a 55063 c 53844 g 95007 t 42 others
ORIGIN

alignment_scores:
  Quality: 38.00      Length: 10
  Ratio: 3.800       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 70.000

alignment_block:
US-08-653-294-15 x AC006858/rev ..
Align seg 1/1 to reverse of: AC006858 from: 1 to: 299719

1 TyArgLeuLeuIleArgLeuAspGluArg 10
|||||:|||||:|||||:|||||:|||||:
163290 TACAGGTATTGATTCGTATGGAGACAGA 163261

seq_name: gb_p12:AF026087

seq_documentation_block:
LOCUS AF026087 2253 bp DNA PLN 13-NOV-1997
DEFINITION Schizosaccharomyces pombe Sati gene, partial cds.
ACCESSION AF026087

```



```

VERSION      AF026087.1 GI:2612954
KEYWORDS
SOURCE       fission yeast.
ORGANISM     Schizosaccharomyces pombe
              Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
              Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE    1 (bases 1 to 2253)
AUTHORS      Seth-Smith,H., Kominami,K. and Toda,T.
TITLE        Direct Submission
JOURNAL      Submitted (19-SEP-1997) Cell Regulation, Imperial Cancer Research
              Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
FEATURES
  source     1..2253
              Location/Qualifiers
              /organism="Schizosaccharomyces pombe"
              /strain="972h"
              /db_xref="taxon:4896"
              <898..>2223
              /gene="sat1"
              <898..>2223
              /gene="sat1"
              /note="8-111"
              <898..2223
              /gene="sat1"
              /function="restores G1 arrest under starvation conditions
              to sat1- mutant"
              /codon_start=1
              /product="Sat1"
              /protein_id="AAB84238.1"
              /db_xref="GI:2612955"
              /translation="LITDKEQOENAIPIVLLRPDILGVDITLSPGEEKCFRLKRRIP
              LRTAPYKGLAFRQFHALVIGVSATNNACVEHEFTITVPVQOOPELPGLKNSGNR
              QSKLFADSLAFIETDAAEIEVSAEKARQFGKNGIRLSODAQEEKQKKYEQHI
              KKLDDSENGEGGQDHDGMRSEINEEVLDEDEKIALKKLESQPMATRFETIS
              KQVLCRLLSKRYRTGEMMTELEGLDAMVQVHMOLESVERIVPDIRNSVQTE
              RATKYKCRITGCTVLENNMSMVPECPFEFTQOQGVHEFLVELLRVANKRE
              MQGPAPRSEEVQSRSRYSRYSEEVORSEGLSEVSEEEKKGSGQHKRMASL
              TNSREMOIHAPITFAETIQCVLPRIKRVVMEDGMEGGMGQPV"
  BASE COUNT      750 a 342 c 595 g 566 t
  ORIGIN
alignment_scores
  Quality:      37.00      Length:      9
  Ratio:        4.625     Gaps:      0
  Percent Similarity: 88.889 Percent Identity: 88.889
alignment_block
  US-08-653-294-15 x AF026087/rev ..
  Align seg 1/1 to reverse of: AF026087 from: 1 to: 2253
      1 TyrArgLeuLeuIleArgLeuAspGlu 9
      |||||
      976 TATGACTCCCTAATATATCTGGACGAG 950
seq_name: gb_pri:HUMM6PR
seq documentation block:
LOCUS      HUMM6PR      2428 bp      mRNA      PRI      07-JAN-1995
DEFINITION Human cation-dependent mannose 6-phosphate-specific receptor mRNA,
              complete cds.
ACCESSION  M16985 J02937 M19258 M19259
VERSION    M16985.1 GI:187282
KEYWORDS   mannose 6-phosphate-specific receptor.
SOURCE     Human placenta, cDNA to mRNA (library of Clontech), clones
              Pl4a.291.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2428)
AUTHORS    Pohlmann,R., Nagel,G., Schmidt,B., Stein,M., Lorkowski,G.,
              Krentler,C., Cully,J., Meyer,H.E., Grzeschik,K.-H., Mersmann,G.,
              Hasilik,A. and von Figura,K.
TITLE      Cloning of a cDNA encoding the human cation-dependent mannose
              6-phosphate-specific receptor
              Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5575-5579 (1987)
              87289647
COMMENT    The protein is a presumptive prepropeptide, though the cleavage
              site for the mature peptide was not determined.
              Draft entry and computer readable copy of sequence [1] kindly
              provided by K.von Figura, 01-OCT-1987.
FEATURES
  Location/Qualifiers
  1..2428
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /map="12"
  <1..2428
  /note="mps receptor mRNA"
  /gene="M6PR"
  /note="mannose 6-phosphate-specific receptor protein
  signal peptide"
  146..979
  /gene="M6PR"
  146..979
  /gene="M6PR"
  /note="mannose 6-phosphate-specific receptor protein
  precursor"
  /codon_start=1
  /db_xref="GDB:G00-120-162"
  /protein_id="AAA59542.1"
  /db_xref="GI:307147"
  /translation="MFPYSCWRTGLLLLLAVAVRESQTEEKTCDLVGEKGKSEK
  ELALVKRLKPLNFKSFESTVGGSDTYIIFRCVREAGNHTSGAGLVQINKSNGKETV
  VGLNETHIFNGSNMIMLYKGGDEYDHCGRQRRVVMISCNHRTLADNFNPYSEE
  RGVQDCFLFEMDSLLACSPSISHLVSGSILLVTFASLVAVYVVGGLYQRLVYVGA
  GMEQPHLAFWQDLGNLVADGCDFCVRSKPRNPAAYRGVGGDQLGESEERDHLPL
  M"
  mat_peptide 224..976
  /gene="M6PR"
  /note="mannose 6-phosphate-specific receptor protein"
  BASE COUNT      618 a 534 c 576 g 700 t
  ORIGIN
alignment_scores
  Quality:      37.00      Length:      10
  Ratio:        3.700     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 60.000
alignment_block
  US-08-653-294-15 x HUMM6PR/rev ..
  Align seg 1/1 to reverse of: HUMM6PR from: 1 to: 2428
      1 TyrArgLeuLeuIleArgLeuAspGluarg 10
      |||||
      1832 TACCAGCTTTTACTACAAATGGATGAAGA 1803

```

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-15 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Feb 8, 2000 1:28 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=Cgmi\_1/USPTO\_spool/US08653294/runat\_04022000\_160701\_15807/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -QFMT=fastp -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cgi -LISTN=45 -DOCALIGN=200 -THR\_SCORE=pct  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

Search information block:

Query: US-08-653-294-15

Query length: 10

Database: N\_Geneseq\_36:\*

Database sequences: 311585

Database length: 125096042

Search time (sec): 590.520000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:V74662	+	38.00	120.71	40.09	1816	!
N_Geneseq_36:X06749	+	37.00	99.03	646.49	14024	!
N_Geneseq_36:Q25594	+	36.00	110.44	149.75	2485	!
N_Geneseq_36:V49983	+	36.00	109.92	160.02	2836	!
N_Geneseq_36:V2120910	-	35.00	73.45	1.7e+04	110000	!
N_Geneseq_36:X01698	+	33.00	111.56	129.68	603	!
N_Geneseq_36:X01701	+	33.00	101.56	129.68	603	!
N_Geneseq_36:X03749	+	33.00	103.58	360.67	1498	!
N_Geneseq_36:X03754	+	33.00	102.30	425.14	1734	!
N_Geneseq_36:X03748	+	33.00	102.30	425.14	1734	!
N_Geneseq_36:X03767	+	33.00	87.03	3.0e+03	9898	!
N_Geneseq_36:T33183	+	33.00	86.78	3.1e+03	10190	!
N_Geneseq_36:X02058	+	33.00	86.51	3.2e+03	10502	!
N_Geneseq_36:Q87446	+	32.50	97.20	817.95	2504	!
N_Geneseq_36:X02081	+	32.00	119.99	43.98	150	!
N_Geneseq_36:V75507	+	32.00	114.71	86.58	274	!
N_Geneseq_36:T84095	+	32.00	103.08	384.42	1032	!
N_Geneseq_36:T29262	+	32.00	100.70	521.65	1354	!
N_Geneseq_36:T29774	+	32.00	97.11	826.96	2040	!
N_Geneseq_36:N90712	+	32.00	96.86	854.35	2100	!
N_Geneseq_36:Q14809	+	32.00	94.33	1.2e+03	2802	!
N_Geneseq_36:T15599	+	32.00	94.09	1.2e+03	2880	!
N_Geneseq_36:Q14810	+	32.00	94.03	1.2e+03	2901	!
N_Geneseq_36:V36211	+	32.00	90.67	1.9e+03	4256	!
N_Geneseq_36:V83003	+	32.00	90.67	1.9e+03	4256	!
N_Geneseq_36:X26856	+	32.00	90.15	2.0e+03	4516	!
N_Geneseq_36:X13352	+	32.00	87.25	2.9e+03	6285	!
N_Geneseq_36:X03560	+	32.00	84.64	4.1e+03	8467	!
N_Geneseq_36:T80415	+	32.00	80.24	7.2e+03	13987	!
N_Geneseq_36:N50107	+	32.00	79.54	7.8e+03	15155	!
N_Geneseq_36:T78508	+	32.00	70.12	2.6e+04	44377	!
N_Geneseq_36:T80414	+	32.00	70.12	2.6e+04	44377	!
N_Geneseq_36:V15570	+	31.00	128.26	15.23	38	!
N_Geneseq_36:V36099	+	31.00	104.39	325.29	579	!
N_Geneseq_36:V45819	+	31.00	98.54	688.40	1128	!
N_Geneseq_36:T79695	+	31.00	98.40	701.45	1147	!
N_Geneseq_36:T73474	+	31.00	94.60	1.1e+03	1769	!
N_Geneseq_36:V81791	+	31.00	93.11	1.4e+03	2097	!
N_Geneseq_36:T31725	+	31.00	92.97	1.4e+03	2129	!
N_Geneseq_36:V84064	+	31.00	92.54	1.5e+03	2238	!
N_Geneseq_36:T58392	+	31.00	92.53	1.5e+03	2239	!

N\_Geneseq\_36:V41261 - 31.00 91.88 1.6e+03 2413 ! Chlamydomonas reinhardtii R  
N\_Geneseq\_36:Q36520 - 31.00 91.33 1.7e+03 2569 ! Sequence of the tapetum-spe  
N\_Geneseq\_36:T42791 - 31.00 90.59 1.9e+03 2794 ! Scytalidium catalase gene.  
N\_Geneseq\_36:V59714 + 31.00 87.92 2.7e+03 3791 ! Human secreted protein gene

seq\_name: N\_Geneseq\_36:V74662

seq\_documentation\_block:

ID V74662 standard; DNA: 1816 BP.

AC V74662;

DE 16-MAR-1999 (first entry)

DT Staphylococcus aureus contig SEQ ID #351.

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;

KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;

KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

FH Key Location/Qualifiers

FT misc\_feature 241..300

FT /\*tag= a

FT /note= "these bases represent a line of missing text in

the sequence listing in the specification. They

are included to maintain the nucleotide numbering

given in the specification for this DNA sequence"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

alignment\_scores:

Quality: 38.00 Length: 9

Ratio: 4.750 Gaps: 0

Percent Similarity: 88.889 Percent Identity: 88.889

alignment\_block:

US-08-653-294-15 x V74662/rev

Align seg 1/1 to reverse of: V74662 from: 1 to: 1816

1 TyrArgLeuLeuLeuArgLeuAspGlu 9

1387 TATCGTTTACTTATACATTACCGAA 1361

seq\_name: N\_Geneseq\_36:X06749

```

seq_documentation_block:
ID X06749 standard; DNA: 14024 BP.
AC X06749:
DT 26-APR-1999 (first entry)
DE E. coli O157 antigen gene cluster.
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
  wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
OS Escherichia coli.
FH Key Location/Qualifiers
FT CDS 79..861
FT FT /*tag= a
FT FT /product= sugar_transferase
FT FT /note= "wbdN gene (ORF1), this region is
  specifically claimed in Claim 6; encodes W88312"
FT CDS 858..2042
FT FT /*tag= b
FT FT /product= O_antigen-polymerase
FT FT /note= "wzy gene (ORF2), this region is
  specifically claimed in Claim 6; encodes W88313"
FT CDS 2011..2757
FT FT /*tag= c
FT FT /product= sugar_transferase
FT FT /note= "wbdO gene (ORF3), this region is
  specifically claimed in Claim 6; encodes W88314"
FT CDS 2744..4135
FT FT /*tag= d
FT FT /product= O_antigen_flippase
FT FT /note= "wzx gene (ORF4), this region is
  specifically claimed in Claim 6; encodes W88315"
FT CDS 4132..5232
FT FT /*tag= e
FT FT /product= perosamine_synthetase
FT FT /note= "per gene (ORF5); encodes W88316"
FT CDS 5257..6471
FT FT /*tag= f
FT FT /product= sugar_transferase
FT FT /note= "wbdP gene (ORF6), this region is
  specifically claimed in Claim 6; encodes W88317"
FT CDS 6491..7609
FT FT /*tag= g
FT FT /note= "ORF7; encodes W88318"
FT CDS 7606..8578
FT FT /*tag= h
FT FT /note= "ORF8; encodes W88319"
FT CDS 8580..9089
FT FT /*tag= i
FT FT /note= "ORF9; encodes W88320"
FT CDS 9071..10519
FT FT /*tag= j
FT FT /note= "manC gene (ORF10); encodes W88321"
FT CDS 10523..11893
FT FT /*tag= k
FT FT /note= "manD gene (ORF11); encodes W88322"
FT FT /tag= l
FT FT /note= "H-repeat, may play a role in gene cluster
  assembly"
FT CDS 13156..13821
FT FT /*tag= m
FT FT /product= N-acetyl_transferase
FT FT /note= "wbdR gene (ORF12), this region is
  specifically claimed in Claim 6; encodes W88323"
FT PN W09850531-A1.
FT PD 12-NOV-1998.
FT PF 01-MAY-1998; AU0315.
FT PR 22-JUL-1997; AU-008162.
FT PR 01-MAY-1997; AU-006545.
FT PA (UNSY ) UNIV SYDNEY.
FT PI Reeves PR, Wang L;
DR WPI; 99-059669/05;
DR DR P-PSDB; W88312, W88313, W88314, W88315, W88316, W88317, W88318, W88319,
  W88320, W88321, W88322, W88323.

```

PT Nucleic acid molecules specific for bacterial polysaccharide  
PT antigens - useful for detecting specific strains in, e.g. food,  
PT faeces or patient samples  
PT Claim 6; Fig 6; 165pp; English.  
CC This is the nucleotide sequence of a gene cluster involved in  
CC the biosynthesis of the Escherichia coli O157 O antigen. The  
CC gene cluster was obtained by PCR amplification of E. coli  
CC O157:H7 strain (C664-1992) DNA using primers (see X06752-63) based  
CC on O antigen gene sequences. 12 open reading frames within the  
CC gene cluster were identified on the basis of homology to known  
CC sequences. The genes encode proteins (see W88312-23) involved in  
CC synthesis of sugars present in the polysaccharide antigen, and in  
CC the transport or processing of polysaccharide or oligosaccharide  
CC units. The use of nucleic acid molecules derived from particular  
CC assembly and transport genes, particularly wbd (transferase), wzx  
CC (flippase) and wzy (polymerase) genes, within O antigen gene  
CC clusters improves the specificity of methods for the detection and  
CC identification of O antigens, e.g. in tests of food- or faecal-  
CC derived samples, or samples from patients. The O antigen is a  
CC major virulence factor of enteropathogenic E. coli strains that  
CC cause diarrhoea and haemorrhagic colitis.  
SQ Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T;

alignment\_scores:  
Quality: 37.00 Length: 9  
Ratio: 4.111 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 66.667

alignment\_block:  
US-08-653-294-15 x X06749/rev ..  
Align seg 1/1 to reverse of: X06749 from: 1 to: 14024

1 TyrArgLeuLeuIleArgLeuAspGlu 9  
6717 TTTAGAAATCTAGTGGTTAGATGAA 6691

seq\_name: N\_Geneseq\_36:Q25594

seq\_documentation\_block:  
ID Q25594 standard; cDNA; 2485 BP.  
AC Q25594:  
DT 03-DEC-1992 (first entry)  
DE Rabbit Gnt I cDNA clone.  
KW High mannose glycoprotein; UDP-N-acetyl-glucosamine;  
KW alpha-3-D-mannoside; beta-1,2-N-acetyl-glucosaminyl transferase I;  
KW N-glycans; ss.  
OS Oryctolagus cuniculus.  
FH Key Location/Qualifiers  
FT cds 50..1393  
FT /\*tag= a  
FT FT /product= Gnt\_I  
FT poly\_a\_signal 2435..2440  
FT FT /\*tag= b  
FT PN W09209694-A.  
FT PD 11-JUN-1992.  
FT PF 29-NOV-1991; CA0417.  
FT PR 30-NOV-1990; US-620098.  
FT PA (HSCR-) HSC RES & DEV LP.  
FT PI Sarkar M, Schachter H;  
FT DR WPI; 92-217073/26.  
FT DR P-PSDB; R24779.  
FT PT Human and rabbit DNA sequences encoding GNT I enzymes - for  
FT converting mannose to hybrid and complex N-glycan(s)  
FT Claim 3; Fig 4; 57pp; English.  
CC Rabbit liver Gnt I was purified and digested with pepsin, then  
CC trypsin. The peptide fragments were sequenced and used to design  
CC degenerate PCR primers (see Q30220-5). cDNA was prepared from total  
CC RNA from rabbit liver. PCR was carried out on the cDNA preparation.  
CC One of the two PCR products was cloned into the SmaI site of pGEM7z  
CC for sequencing and then used as a riboprobe. The riboprobe was used  
CC to screen a rabbit liver cDNA library in lambda gt10. The largest

CC insert in a positive clone was 1.6kb. An 80bp riboprobe was  
 CC prepared from the 5'-terminal of the 1.6kb insert and used to  
 CC rescreen the library. The largest cDNA insert was cloned into  
 CC pGEM-7z to obtain pGEM-7z-rcnpt1. The full-length rabbit Gnt I  
 CC coding sequence was eventually obtained from overlapping clones.  
 CC See also Q25595.  
 SQ Sequence 2485 BP; 549 A; 623 C; 644 G; 669 T;

alignment\_scores:  
 Quality: 36.00 Length: 9  
 Ratio: 4.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 66.667

alignment\_block:  
 US-08-653-294-15 x Q25594/rev ..  
 Align seg 1/1 to reverse of: Q25594 from: 1 to: 2485

1 TyrArgLeuLeuIleArgLeuAspGlu 9  
 |||||:|||||:|||||:|||||  
 1072 TACAAACTGCTGGTTCAGCTTGATGAA 1046

seq\_name: N\_Geneseq\_36:V49983

seq\_documentation\_block:  
 ID V49983 standard; DNA; 2636 BP.  
 AC V49983;  
 DT 10-NOV-1998 (first entry)  
 DE Floral organ-specific chitinase gene with promoter sequence.  
 KW Floral-organ specific gene; chitinase; promoter; Oryza sativa; rice;  
 KW flower part; monocotyledonous plant; bacterial resistance; ds.  
 OS Oryza sativa.  
 FH Key Location/Qualifiers  
 FT Promoter 1..1234

FT /\*tag= a  
 FT /note= "floral-organ specific promoter sequence"  
 FT 1235..2505  
 FT /\*tag= b  
 FT /product= "floral-organ-specific chitinase"  
 FT /note= "contains introns"  
 FT 1235..1691  
 FT /\*tag= c  
 FT /number= 1  
 FT 1692..1776  
 FT /\*tag= d  
 FT /number= 1  
 FT 1777..1909  
 FT /\*tag= e  
 FT /number= 2  
 FT 1910..2108  
 FT /\*tag= f  
 FT /number= 2  
 FT 2109..2502  
 FT /\*tag= g  
 FT /number= 3

WO9829542-A1.  
 PD 09-JUL-1998.  
 PF 26-DEC-1997; J04892.  
 PR 27-DEC-1996; JP-349505.  
 PA (NISR ) JAPAN TOBACCO INC.  
 PI Inoue T, Ito T, Saito H, Takakura Y;  
 DR WPI: 98-388123/33.  
 DR P-PSDB: W64776

PT Promoter sequence associated with chitinase, specific to floral  
 PT organs - allows expression of genes in the flower parts of  
 PT monocotyledonous plants  
 PS Claim 1; Pages 40-41; 67pp; Japanese.  
 CC This represents a floral organ-specific chitinase encoding gene sequence  
 CC along with a floral organ-specific promoter sequence. The promoter  
 CC derived from rice associated with chitinase is specific to floral organs  
 CC and allows expression of genes in the flower parts of monocotyledonous  
 CC plants. The expression of desired genes specifically in the floral parts

CC of monocotyledonous plants, allows modifications to be made in the  
 CC structure or function of these parts allowing the introduction of desired  
 CC characteristics such as increased resistance to bacterial attack.  
 SQ Sequence 2636 BP; 657 A; 722 C; 696 G; 561 T;

alignment\_scores:  
 Quality: 36.00 Length: 9  
 Ratio: 4.500 Gaps: 0  
 Percent Similarity: 88.889 Percent Identity: 77.778

alignment\_block:  
 US-08-653-294-15 x V49983/rev ..  
 Align seg 1/1 to reverse of: V49983 from: 1 to: 2636

1 TyrArgLeuLeuIleArgLeuAspGlu 9  
 |||||:|||||:|||||:|||||  
 213 TACAGGCGATTAAATAACTAGATGAA 187  
 seq\_name: N\_Geneseq\_36:V21209\_10

seq\_documentation\_block:  
 Continuation (11 of 17) of V21209 from base 1000001 (Methanococcus jannaschii circular  
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP Fragment Name Begin End  
 WP V21209\_00 1 110000  
 WP V21209\_01 100001 210000  
 WP V21209\_02 200001 310000  
 WP V21209\_03 300001 410000  
 WP V21209\_04 400001 510000  
 WP V21209\_05 500001 610000  
 WP V21209\_06 600001 710000  
 WP V21209\_07 700001 810000  
 WP V21209\_08 800001 910000  
 WP V21209\_09 900001 1010000  
 WP V21209\_10 1000001 1110000  
 WP V21209\_11 1100001 1210000  
 WP V21209\_12 1200001 1310000  
 WP V21209\_13 1300001 1410000  
 WP V21209\_14 1400001 1510000  
 WP V21209\_15 1500001 1610000  
 WP V21209\_16 1600001 1664976

alignment\_scores:  
 Quality: 35.00 Length: 10  
 Ratio: 3.889 Gaps: 0  
 Percent Similarity: 90.000 Percent Identity: 60.000

alignment\_block:  
 US-08-653-294-15 x V21209\_10/rev ..  
 Align seg 1/1 to reverse of: V21209\_10 from: 1 to: 110000

1 TyrArgLeuLeuIleArgLeuAspGluArg 10  
 |||||:|||||:|||||:|||||  
 97039 TACAGGAGATTGCTGAGTACCTAGACGAGG 97010

seq\_name: N\_Geneseq\_36:X01698

seq\_documentation\_block:  
 ID X01698 standard; DNA; 603 BP.  
 AC X01698;  
 DT 08-JUN-1999 (first entry)  
 DE Human anti-angiogenic hPRL Met-1cys199 antisense DNA.  
 KW Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis;  
 KW growth hormone; hGH; hGH-V; capillary endothelial cell proliferation;  
 KW placental vascularisation; pregnancy; treatment; angiogenic disease;  
 KW tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation;  
 KW arthritis; atherosclerotic plaques; corneal graft neovascularisation;  
 KW wound healing; proliferative retinopathy; macular degeneration; trachoma;  
 KW granulation; glaucoma; ocular; uveitis; fracture; Osler-Weber syndrome;



KW maize; wheat; tryptophan content; nutritional value; ss.  
 OS Oryza sativa.  
 FH Key Location/Qualifiers  
 FT CDS 1102..1498  
 FT /\*tag= a  
 FT /product= "ASA synthase"  
 FT /note= "Partial coding sequence. Incomplete exon 2"  
 FT exon 1102..1233  
 FT /\*tag= b  
 FT /number= 1  
 FT intron 1234..1318  
 FT /\*tag= c  
 FT /number= 1  
 FT exon 1318..1498  
 FT /\*tag= d  
 FT /number= 2  
 FT /note= "partial exon 2 sequence"

PN WO9911800-A1.  
 PD 11-MAR-1999.  
 PR 31-AUG-1998; J03883.  
 PF 29-AUG-1997; JP-235049.  
 PA (HOKK ) HOKKO CHEM IND CO LTD.  
 PI Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;  
 DR WPI; 99-228982/19.  
 DR P-PSDB; W93811.  
 PT DNA encodes 'a'-subunit of first isozyme of rice anthranilate  
 PT synthase - used for improving tryptophan production and  
 PT nutritional value of crops, e.g. rice, maize or wheat  
 PS Example 5; Page 131-133; 152pp; Japanese.  
 CC This invention describes a novel rice anthranilate synthase first  
 CC isozyme alpha-subunit. The encoding DNA can be used to produce  
 CC transformant plants and seeds, of e.g. rice, maize or wheat, with  
 CC enhanced tryptophan content and nutritional value of the crops.  
 CC enhanced tryptophan content; nutritional value; ss.  
 SQ Sequence 1498 BP; 343 A; 362 C; 404 G; 389 T;

alignment\_scores:  
 Quality: 33.00 Length: 10  
 Ratio: 4.125 Gaps: 0  
 Percent Similarity: 80.000 Percent Identity: 60.000

alignment\_block:

US-08-653-294-15 x X23749 ..

Align seg 1/1 to: X23749 from: 1 to: 1498

1 TyrArgLeuLeuIleArgLeuAspGluArg 10  
 ||||| |||:|||| |:::||||  
 1427 TACCGCTGCTCGTCAGGAGGACGACCGC 1456

seq\_name: N\_Geneseq\_36:X23754

seq\_documentation\_block:

ID X23754 standard; DNA; 1734 BP.

AC X23754;

DT 25-JUN-1999 (first entry)

DE Rice anthranilate synthase first isozyme alpha-subunit DNA variant.

KW Anthranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed;

KW maize; wheat; tryptophan content; nutritional value; ss.

OS Oryza sativa.

FH Key Location/Qualifiers

FT CDS 1..1734

FT /\*tag= a

FT /product= "ASA synthase"

FT WO9911800-A1.

PN 11-MAR-1999.

PR 31-AUG-1998; J03883.

PF 29-AUG-1997; JP-235049.

PA (HOKK ) HOKKO CHEM IND CO LTD.

PI (HOKK ) JAPAN MIN AGRIC FORESTRY & FISHERIES.

PI Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;

DR WPI; 99-228982/19.

DR P-PSDB; W93815.  
 PT DNA encodes 'a'-subunit of first isozyme of rice anthranilate  
 PT synthase - used for improving tryptophan production and  
 PT nutritional value of crops, e.g. rice, maize or wheat  
 PS Claim 8; Page 142-145; 152pp; Japanese.  
 CC This invention describes a novel rice anthranilate synthase first  
 CC isozyme alpha-subunit. The encoding DNA can be used to produce  
 CC transformant plants and seeds, of e.g. rice, maize or wheat, with  
 CC enhanced tryptophan content and nutritional value of the crops.  
 CC enhanced tryptophan content; nutritional value; ss.  
 SQ Sequence 1734 BP; 422 A; 389 C; 505 G; 418 T;

alignment\_scores:  
 Quality: 33.00 Length: 10  
 Ratio: 4.125 Gaps: 0  
 Percent Similarity: 80.000 Percent Identity: 60.000

alignment\_block:

US-08-653-294-15 x X23754 ..

Align seg 1/1 to: X23754 from: 1 to: 1734

1 TyrArgLeuLeuIleArgLeuAspGluArg 10

||||| |||:|||| |:::||||

241 TACCGCTGCTCGTCAGGAGGACGACCGC 270

seq\_name: N\_Geneseq\_36:X23748

seq\_documentation\_block:

ID X23748 standard; DNA; 1734 BP.

AC X23748;

DT 25-JUN-1999 (first entry)

DE Rice anthranilate synthase first isozyme alpha-subunit DNA.

KW Anthranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed;

KW maize; wheat; tryptophan content; nutritional value; ss.

OS Oryza sativa.

FH Key Location/Qualifiers

FT CDS 1..1734

FT /\*tag= a

FT /product= "ASA synthase"

PN WO9911800-A1.

PD 11-MAR-1999.

PR 31-AUG-1998; J03883.

PF 29-AUG-1997; JP-235049.

PA (HOKK ) HOKKO CHEM IND CO LTD.

PA (HOKK ) JAPAN MIN AGRIC FORESTRY & FISHERIES.

PI Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;

DR WPI; 99-228982/19.

DR P-PSDB; W93810.

PT DNA encodes 'a'-subunit of first isozyme of rice anthranilate

PT synthase - used for improving tryptophan production and

PT nutritional value of crops, e.g. rice, maize or wheat

PS Claim 2; Page 125-128; 152pp; Japanese.

CC This invention describes a novel rice anthranilate synthase first.

CC isozyme alpha-subunit. The encoding DNA can be used to produce

CC transformant plants and seeds, of e.g. rice, maize or wheat, with

CC enhanced tryptophan content and nutritional value of the crops.

SQ Sequence 1734 BP; 421 A; 389 C; 506 G; 418 T;

alignment\_scores:  
 Quality: 33.00 Length: 10  
 Ratio: 4.125 Gaps: 0  
 Percent Similarity: 80.000 Percent Identity: 60.000

alignment\_block:

US-08-653-294-15 x X23748 ..

Align seg 1/1 to: X23748 from: 1 to: 1734

1 TyrArgLeuLeuIleArgLeuAspGluArg 10

||||| |||:|||| |:::||||

241 TACCGCTGCTCGTCAGGAGGACGACCGC 270

seq\_name: N\_Geneseq\_36:V20767

seq\_documentation\_block:

ID V20767 standard; DNA; 9898 BP.  
AC V20767;  
DT 15-JUL-1998 (first entry)  
DE Human OCIF genome DNA-2.  
KW Human; OCIF; genome; osteoclast; antipyretic; osteoporosis;  
KW rheumatism; multiple sclerosis; ds.  
OS Homo sapiens.  
FH Key  
CDS

Location/Qualifiers

FT 130..9057  
FT /\*tag= a  
FT /note= "contains introns"  
FT 130..499  
FT exon  
FT /\*tag= b  
FT /number= 1  
FT intron  
FT 500..4503  
FT /\*tag= c  
FT /number= 1  
FT exon  
FT 4504..4695  
FT /\*tag= d  
FT /number= 2  
FT intron  
FT 4696..6715  
FT /\*tag= e  
FT /number= 2  
FT exon  
FT 6716..6940  
FT /\*tag= f  
FT /number= 3  
FT intron  
FT 6941..8668  
FT /\*tag= g  
FT /number= 3  
FT exon  
FT 8669..9057  
FT /\*tag= h  
FT /number= 4

WO9807840-AL.

PD 26-FEB-1998.  
PF 19-AUG-1997; J02859.  
PR 19-AUG-1996; JP-235928.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Morinaga T, Nakagawa N, Yasuda H;  
DR WPI: 98-169150/15.  
DR P-PSDB: W53238.  
PT Inhibition of osteoclast formation and/or antipyretic activity -  
PT useful for, e.g. treating osteoporosis, rheumatism and multiple  
PT sclerosis  
PS Claim 1; Page 16-26; 36pp; English.  
CC The present sequence represents human OCIF genome DNA which is  
CC specifically claimed in the present invention. The present invention  
CC describes: (1) a method of inhibiting the formation of osteoclasts  
CC and/or antipyretic proteins, which have the following characteristics:  
CC (i) MW determined by SDS-PAGE of approximately 60 kDa under reducing  
CC conditions, and 60 kDa and 120 kDa under non-reducing conditions; (ii)  
CC 401 aa sequence (III) (see W53239); (iii) affinity to cation exchangers  
CC and heparin; (iv) its osteoclast formation inhibiting activity reduced  
CC by heating at 70 degrees Celsius for 10 minutes or 56 degrees Celsius  
CC for 30 minutes, and (v) its antipyretic activity reduced by heating at  
CC 90 degrees Celsius for 10 minutes, and (2) a method for preparing the  
CC above proteins. The proteins are useful for, e.g. treatment and  
CC prevention of osteoporosis, rheumatism or multiple sclerosis, and also  
CC as antigens for immunological diagnosis of these diseases and disorders.  
SQ Sequence 9898 BP; 3110 A; 1835 C; 1871 G; 3082 T;

alignment\_scores:

Quality: 33.00 Length: 10  
Ratio: 3.300 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 60.000

alignment\_block:

US-08-653-294-15 x V20767 ..

Align seg 1/1 to: V20767 from: 1 to: 9898

1 TyrArgLeuLeuIleArgLeuAspGluArg 10

||||:|||||  
2648 TATGAACCTTCTCATTAGGATGCAGGAG 2677

seq\_name: N\_Geneseq\_36:T33183

seq\_documentation\_block:

ID T33183 standard; DNA; 10190 BP.  
AC T33183;  
DT 23-APR-1997 (first entry)  
DE Fragment of human OCIF genomic DNA-2.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis; ds.  
OS Homo sapiens.  
FH Key  
CDS

Location/Qualifiers

FT 130..499  
FT /\*tag= a  
FT /note= "encodes residues 11 to 112 of OCIF, see R99925"  
FT 500..4503  
FT /\*tag= b  
FT exon  
FT 4504..4695  
FT /\*tag= c  
FT /note= "encodes residues 113 to 176 of OCIF, see R99925"  
FT 4696..6715  
FT /\*tag= d  
FT exon  
FT 6716..6940  
FT /\*tag= e  
FT /note= "encodes residues 177 to 251 of OCIF, see R99925"  
FT 6941..8960  
FT intron  
FT 8961..9349  
FT /\*tag= f  
FT exon  
FT 9350..9500  
FT /\*tag= g  
FT /note= "encodes residues 252 to 380 of OCIF, see R99925"

WO9626217-AL.

PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 9; Page 152-161; 183pp; Japanese.  
CC The sequences given in T33182-83 represent fragment of the genomic  
CC DNA encoding the full length osteoclastogenesis inhibitory factor  
CC (OCIF) of the invention. The OCIF of the invention has a  
CC molecular weight by SDS-PAGE of 60 kD under reducing conditions and  
CC 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 10190 BP; 3184 A; 1920 C; 1899 G; 3187 T;

alignment\_scores:

Quality: 33.00 Length: 10  
Ratio: 3.300 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 60.000

alignment\_block:

US-08-653-294-15 x T33183 ..  
Align seg 1/1 to: T33183 from: 1 to: 10190

1 TyrArgLeuLeuIleArgLeuAspGluArg 10

||||:|||||  
2648 TATGAACCTTCTCATTAGGATGCAGGAG 2677



```

seq_name: N_Geneseq_36:X20258
seq_documentation_block:
ID X20258 standard; DNA; 10502 BP.
AC X20258;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #11.
DE Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN W09858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1: Page 908-914; 1128bp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
CC Sequence 10502 BP; 3939 A; 1402 C; 1401 G; 3760 T;
SQ Sequence 10502 BP; 3939 A; 1402 C; 1401 G; 3760 T;

alignment_scores:
Quality: 33.00 Length: 8
Ratio: 4.125 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000

alignment_block:
US-08-653-294-15 x X20258/rev ..
Align seg 1/1 to reverse of: X20258 from: 1 to: 10502

1 TyrArgLeuLeulleAtgLeuAsp 8
|||||:|||||:|||||:|||||
8697 TAGAATACCTCTTCGGCTTGAT 8674

seq_name: N_Geneseq_36:Q87446
seq_documentation_block:
ID Q87446 standard; CDNA; 2504 BP.
AC Q87446;
DT 21-NOV-1995 (first entry)
DE Tribolium semaphorin I cDNA.
DE Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
KW varicella major virus; smallpox; semaphorin receptor binding activity;
KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection; ds.
OS Tribolium sp.
FH Key Location/Qualifiers
FT cds 355..2493
FT /*tag= a
FT /*product= semaphorin I
FT W09507706-A.
PN 23-MAR-1995.
PF 13-SEP-1994; U10151.

seq_name: N_Geneseq_36:X12081
seq_documentation_block:
ID X12081 standard; DNA; 150 BP.
AC X12081;
DT 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment EST91495b.
DE Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI: 98-286974/25
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1: Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos

```

```

PR 13-SEP-1993; US-121713.
PA (REGC ) UNIV CALIFORNIA.
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;
PI O'Connor T;
DR WPI: 95-131177/17.
DR P-PSDB; R71384.
PT New class of semaphorin peptide(s) and polypeptide(s) - are
PT potent modulators of nerve cell growth and regeneration
PS Example 2: Page 85-88; 101pp; English.
CC The sequence of the cDNA encoding the beetle Tribolium semaphorin I
CC protein. This sequence was isolated by PCR using Tribolium genomic DNA.
CC The proteins encoded by the grasshopper semaphorin I (Q87441), human
CC semaphorin III (Q87442), vaccinia virus semaphorin IV (Q87443), Drosophila
CC semaphorin I and II (Q87444-5), Tribolium semaphorin I or varicella major
CC (smallpox) virus semaphorin IV (Q87447) genes were used to generate a
CC series of peptides (R70370-R70418), which retain semaphorin receptor
CC binding activity. The semaphorin derived or semaphorin receptor derived
CC peptides are potent modulators of nerve cell growth, immune
CC responsiveness and viral pathogenesis. They can be used in diagnosis and
CC treatment of neurological disease and neuro-regeneration, immune
CC modulation and diagnosis and treatment of viral and oncological infection
CC and diseases.
SQ Sequence 2504 BP; 681 A; 577 C; 614 G; 632 T;

```

```

alignment_scores:
Quality: 32.50 Length: 13
Ratio: 3.250 Gaps: 1
Percent Similarity: 76.923 Percent Identity: 69.231

```

```

alignment_block:
US-08-653-294-15 x Q87446/rev ..
Align seg 1/1 to reverse of: Q87446 from: 1 to: 2504

```

```

1 TyrArgLeuLeulle.....ArgLeuAspGluArg 10
|||||:|||||:|||||:|||||
544 TACCGCCTACTATATCGAGTTCGTCGTGATTCAAGA 506

```

```
seq_name: N_Geneseq_36:X12081
```

```
seq_documentation_block:
ID X12081 standard; DNA; 150 BP.

```

```

AC X12081;
DT 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment EST91495b.
DE Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.

```

```

PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI: 98-286974/25

```

```

PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1: Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos

```

CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 150 BP; 38 A; 38 C; 32 G; 41 T;

alignment\_scores:  
Quality: 32.00 Length: 8  
Ratio: 4.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-08-653-294-15 x X12081/rev ..

Align seg 1/1 to reverse of: X12081 from: 1 to: 150

1 TyrArgLeuLeuIleArgLeuasp 8  
||||:|||||:||||:||||  
52 TACAGCTGCTGCTCAGGATGAC 29

OM of: US-08-653-294-15 to: EST:\* out\_format : pfs

Date: Feb 8, 2000 4:03 AM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frameparam.p2n.model -DEV=xip  
-O=/cgl\_1/USPTO.spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.1  
-DB=EST\_QPWT-fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500  
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-15

Query length: 10

Database: EST:\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 8553.360000

score\_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_gss15:AQ652500	-	39.00	139.97	49.60	491	! AQ652500 Sheared DNA-22K17.TF S
gb_est9:AA105773	-	37.00	140.09	48.90	189	! AA105773 ml84g04.r1 Stratagene
gb_est6:N92217	-	37.00	138.75	58.08	222	! N92217 y290d07.r1 Soares_multit
gb_est38:AA1837241	+	37.00	136.96	75.00	275	! AI837241 UI-M-AK0-ade-f-09-0-UI
gb_est8:AA040214	-	37.00	136.66	75.84	285	! AA040214 z103b05.r1 Soares_feta
gb_est9:AA085256	-	37.00	135.49	88.13	338	! AA085256 znl2g08.r1 Stratagene
gb_est5:H87592	-	37.00	135.10	92.73	344	! H87592 yw17g08.r1 Soares_placen
gb_est9:AA106940	-	37.00	134.93	94.75	351	! AA106940 m912h06.r1 Stratagene
gb_est27:AA1413784	+	37.00	133.88	108.36	398	! AT413784 ms91h02.x1 Soares_mous
gb_est15:AA465578	-	37.00	133.80	109.52	402	! AA465578 aa24f06.r1 NCI_CGAP_GC
gb_est38:AA048963	-	37.00	133.70	110.98	407	! AA048963 UI-M-BH1-amo-g-04-0-UI
gb_est5:N28776	-	37.00	133.20	118.28	432	! N28776 yx69g09.r1 Soares_melan
gb_est10:AA137143	+	37.00	133.18	118.57	433	! AA137143 z102f08.r1 Soares_preg
gb_est35:AA1849304	+	37.00	133.16	118.57	434	! AI849304 UI-M-AJ1-aha-b-09-0-UI
gb_est6:NA6203	-	37.00	133.16	118.86	434	! NA6203 yx38f03.r1 Soares_melan
gb_est8:AA036922	-	37.00	132.86	123.55	450	! AA036922 zx32f05.r1 Soares_preg
gb_est7:W88510	-	37.00	132.84	123.84	451	! W88510 zh72g01.r1 Soares_fetal
gb_est14:AA423708	-	37.00	132.73	125.60	457	! AA423708 ve81b03.r1 Soares_mous
gb_est5:H87502	-	37.00	132.34	132.07	479	! H87502 yw14c09.r1 Soares_placen
gb_est9:AA114917	-	37.00	132.32	132.37	480	! AA114917 z103f09.r1 Soares_preg
gb_est15:AA89970	-	37.00	131.82	141.22	510	! AA89970 ab05h01.r1 Stratagene
gb_est6:N53340	-	37.00	131.67	143.89	519	! N53340 yz03f12.r1 Soares_multit
gb_est5:N34201	-	37.00	131.62	144.78	522	! N34201 yz77f03.r1 Soares_melan
gb_est18:AA674758	-	37.00	131.51	146.85	529	! AA674758 vq98c09.r1 Knowles_Sol
gb_est17:AA632641	+	37.00	131.43	148.34	534	! AA632641 np87b04.s1 NCI_CGAP_Th
gb_est9:AA122242	+	37.00	130.98	157.25	564	! AA122242 z101e12.r1 Soares_preg
gb_est7:W22057	-	37.00	129.15	174.89	623	! W22057 63F2 Human retina cDNA
gb_est22:AA1066787	+	37.00	129.98	178.79	636	! AA1066787 ov17b09.x1 NCI_CGAP_GC
gb_est4:AA192536	-	37.00	129.90	180.59	642	! AA192536 xi46c03.x1 NCI_CGAP_Pa
gb_est37:AA1956148	-	37.00	129.78	183.30	651	! AA1956148 wt35e09.x1 NCI_CGAP_Pa
gb_est20:AA897536	-	37.00	129.72	184.80	656	! AA897536 wj62e11.s1 Soares_test
gb_est36:AA188684	-	37.00	129.61	187.51	665	! AA188684 wn34d10.x1 NCI_CGAP_Ga
gb_est38:AA130547	+	37.00	129.42	192.04	680	! AA130547 x52h02.x1 NCI_CGAP_Ga
gb_est39:AAW027634	+	37.00	129.00	202.61	715	! AW027634 wv69g06.x1 Soares_thym
gb_est30:AA1635238	+	37.00	128.78	208.37	734	! AA1635238 tz79e04.x1 NCI_CGAP_GC
gb_est17:AA648974	-	37.00	128.53	215.35	757	! AA648974 ns41e06.s1 NCI_CGAP_GC
gb_est22:AA1005191	+	37.00	128.51	215.96	759	! AA1005191 ou60b10.x1 NCI_CGAP_Bi
gb_gss8:AAQ36625	+	37.00	128.40	219.00	769	! AAQ36625 CJT-HSP-2319N16.TF CIT
gb_est25:AA310218	+	37.00	127.83	235.47	823	! AA310218 q075g01.x1 NCI_CGAP_Ki
gb_gss3:H29703	+	36.00	132.18	134.78	305	! H29703 T16A22TRB TAMU Arabidops
gb_est5:D75467	+	36.00	131.40	148.99	335	! D75467 CELK104H2F Yuji Kohara
gb_est5:D73508	+	36.00	130.80	160.90	360	! D73508 CELK051H2F Yuji Kohara

gb\_est5:D73590 + 36.00 130.80 160.90 360 ! D73590 CELK060G3F Yuji Kohar  
gb\_est5:D74728 + 36.00 130.80 160.90 360 ! D74728 CELK084F3F Yuji Kohar  
gb\_est5:D75423 + 36.00 130.80 160.90 360 ! D75423 CELK104C7F Yuji Kohar

seq\_name: gb\_gss15:AQ652500

seq\_documentation\_block:

LOCUS AQ652500 491 bp DNA GSS 22-JUN-1999  
DEFINITION Sheared DNA-22K17.TF Sheared DNA Trypanosoma brucei genomic clone  
Sheared DNA-22K17, genomic survey sequence.

ACCESSION AQ652500

VERSION AQ652500.1 GI:5145686

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 491)

REFERENCE  
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Sub,E., Malek,J., Fujii,C.,  
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,  
Donelson,J., Fraser,C. and Adams,M.

TITLE  
Determination of clone end sequences from Trypanosoma brucei GUTat  
10.1 sheared DNA library

JOURNAL  
COMMENT Unpublished (1999)

Other\_GSSs: Sheared DNA-22K17.TR

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
DNA library constructed at TIGR. Clones will be available for  
distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tldb/mdb/tldb/>.

Seq primer: M13-Forward

Class: Shotgun

FEATURES

Location/Qualifiers

1..491

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db\_xref="taxon:5691"

/clone\_lib="Sheared DNA-22K17"

/note="Vector: pUC18; Site:1: SmaI; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically

sheared to give a tight size distribution (approx 2 kb).

The v + i method used for the library construction is

described in detail in Smith, H.O. and Venter, J.C.

(Making small insert libraries for whole genome shotgun

sequencing projects. In Genome Sequencing: A Practical

Approach, eds. M. Vaudin and B. Borell, Oxford University

Press, 1999).

BASE COUNT 173 a 122 c 124 g 72 t

ORIGIN

alignment\_scores:

Quality: 39.00 Length: 9

Ratio: 4.333 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 77.778

alignment\_block:

US-08-653-294-15 x AQ652500/rev ..

Align seg 1/1 to reverse of: AQ652500 from: 1 to: 491

1 TyrArgLeuLeuIleArgLeuaspGlu 9

|||||

|||||

|||||

50 TATCGTCTTCTTGTGGCGTGGATAAA 24

```

seq_name: gb_est9:AA105773
seq_documentation_block: 189 bp mRNA EST 04-FEB-1997
LOCUS AA105773
DEFINITION m184904.r1 stratagene mouse kidney (#937315) Mus musculus cDNA
clone IMAGE:518742 5', mRNA sequence.
ACCESSION AA105773
VERSION AA105773.1 GI:1654862
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 189)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilton,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:593450.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1800
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:312590
Putative full length read
vector to vector length is 372
Seq primer: -28mi3 rev1 ET from Amersham
High quality sequence stop: 140.
FEATURES
Location/Qualifiers
1..189
/organism="Mus musculus"
/strain="C57/B16"
/db_xref="taxon:10090"
/clone="IMAGE:518742"
/clone_lib="Stratagene mouse kidney (#937315)"
/sex="females"
/tissue_type="kidney"
/dev_stage="4 weeks"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: Kidney; Vector: pBluescript SK-; Site_1: ECOR1; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'"
BASE COUNT 51 a 60 c 41 g 37 t
ORIGIN
alignment_scores:
Quality: 37.00 Length: 9
Ratio: 4.625 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 88.889
alignment_block:
US-08-653-294-15 x AA105773/rev ..
Align seg 1/1 to reverse of: AA105773 from: 1 to: 189
1 TyrArgLeuLeuIleArgLeuAspGlu 9
147 TACTCTCTCTCTCATCAGACTGGATGAG 121
seq_name: gb_est6:N92217
seq_documentation_block: 222 bp mRNA EST 04-APR-1996
LOCUS N92217
DEFINITION y290d07.r1 Soares_multiple_sclerosis_2NBHMSp Homo sapiens cDNA
clone IMAGE:290317 5' similar to gb:X56253.rnal CARION-DEPENDENT
MANNOS-6-PHOSPHATE RECEPTOR PRECURSOR (HUMAN); contains element THR
repetitive element ;, mRNA sequence.
ACCESSION N92217
VERSION N92217.1 GI:12664526
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 222)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Kucaba,T., Le,M., Lennon,G., Marrs,M., Parsons,J., Ruffin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Sep 21, 1992 this sequence version replaced gi:276322.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1800
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ETP-primer
High quality sequence stop: 47.
FEATURES
Location/Qualifiers
1..222
/organism="Homo sapiens"
/db_xref="GDB:3906157"
/db_xref="taxon:9606"
/clone="IMAGE:290317"
/clone_lib="Soares_multiple_sclerosis_2NBHMSp"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified
polylinker V-type; phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCCATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 52 a 48 c 47 g 75 t
ORIGIN
alignment_scores:
Quality: 37.00 Length: 10
Ratio: 3.700 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 60.000
alignment_block:
US-08-653-294-15 x N92217/rev ..
Align seg 1/1 to reverse of: N92217 from: 1 to: 222
1 TyrArgLeuLeuIleArgLeuAspGluArg 10
125 TACCAGCTTTTACTACAAATGATGAAGA 96

```

Quality:	37.00	Length:	10
Ratio:	3.700	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	60.000

```

alignment_block:
US-08-653-294-15 x AA040214/rev ..
Align seg 1/1 to reverse of: AA040214 from: 1 to: 285

1 TtArgLeuLeuileArgLeuAspGluArg 10
|||||:|||||:|||||:|||||:|||||
216 TACCAGCTTTTACTACAATGATGAAGA 187

seq_name: gb_est9:AA085256

seq_documentation_block:
LOCUS AA085256 328 bp mRNA EST 23-DEC-1997
DEFINITION znl3208.r1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone
IMAGE:547262 5' similar to gb:X56253_rnal CATION-DEPENDENT
MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (HUMAN); mRNA sequence.
ACCESSION AA085256
VERSION AA085256.1 GI:1627314
KEYWORDS EST..
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
7044478
On May 5, 1995 this sequence version replaced gi:797884.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 670 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 198.
Location/Qualifiers
1..328
/organism="Homo sapiens"
/db_xref="GDB:3925452"
/db_xref="taxon:9606"
/clone="IMAGE:547262"
/clone_lib="Stratagene hnt neuron (#937233)"
/dev_stage="hnt neurons"
/lab_host="SOLR (kanamycin resistant)"
/Note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt.
Differentiated, post mitotic hnt neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTGTGTGTGTGTGTGT 3'"

BASE COUNT 68 a 76 c 79 g 92 t 13 others
ORIGIN

seq_name: gb_est5:H87592

seq_documentation_block:
LOCUS H87592 344 bp mRNA EST 21-NOV-1995
DEFINITION yw17g08.r1 Soares_placenta_8to9weeks_2NBHP8to9m Homo sapiens cDNA
clone IMAGE:252542 5' similar to gb:X56253_rnal CATION-DEPENDENT
MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (HUMAN); mRNA sequence.
ACCESSION H87592
VERSION H87592.1 GI:1069171
KEYWORDS EST..
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 344)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:635818.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stop: 330
Source: IMAGE Consortium, LML
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 658 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 330.
Location/Qualifiers
1..344
/organism="Homo sapiens"
/db_xref="GDB:3885608"
/db_xref="taxon:9606"
/clone="IMAGE:252542"
/clone_lib="Soares_placenta_8to9weeks_2NBHP8to9m"
/dev_stage="two placenta: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/Note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCACTGAAGTCGAGCGCGCGGATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 77 a 77 c 86 g 104 t
ORIGIN

alignment_scores:
Quality: 37.00 Length: 10

```

Ratio: 3.700 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 60.000

## alignment\_block:

US-08-653-294-15 x H87592/rev ..

Align seg 1/1 to reverse of: H87592 from: 1 to: 344

1 TyrArgLeuLeuIleArgLeuAspGluArg 10

|||||:|||||:|||||:|||||:|||||

142 TACCAGCTTTTACTACAAATGATGAAGA 113

seq\_name: gb\_est9:AA106940

seq\_documentation\_block: 351 bp mRNA EST 04-FEB-1997  
LOCUS AA106940  
DEFINITION m192h06.r1 Stratagene mouse kidney (#937315) Mus musculus CDNA  
clone IMAGE:519515 5' similar to gb:M22810 Mouse androgen-regulated  
protein mRNA, complete cds (MOUSE);, mRNA sequence.

ACCESSION AA106940

VERSION AA106940.1 GI:1660749

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 351)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Oct 30, 1996 this sequence version replaced gi:1656884.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 448.

## FEATURES

Location/Qualifiers  
1..351  
/organism="Mus musculus"  
/strain="C57/B16"  
/db\_xref="taxon:10090"  
/clone="IMAGE:519515"  
/clone\_lib="Stratagene mouse kidney (#937315)"  
/sex="females"  
/tissue\_type="kidney"  
/dev\_stage="4 weeks"  
/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: kidney; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;  
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 91 a 99 c 81 g 80 t

ORIGIN

## alignment\_scores:

Quality: 37.00 Length: 9  
Ratio: 4.625 Gaps: 0  
Percent Similarity: 88.889 Percent Identity: 88.889

## alignment\_block:

US-08-653-294-15 x AA106940/rev ..

Align seg 1/1 to reverse of: AA106940 from: 1 to: 351

1 TyrArgLeuLeuIleArgLeuAspGlu 9

|||||:|||||:|||||:|||||:|||||

162 TACCTCCTCCTCATCAGACTGATGAG 136

seq\_name: gb\_est27:AI413784

seq\_documentation\_block: 398 bp mRNA EST 09-FEB-1999  
LOCUS AI413784  
DEFINITION me91h02.x1 Soares mouse embryo NbMEL3.5 14.5 Mus musculus CDNA  
clone IMAGE:402963 3', mRNA sequence.

ACCESSION AI413784

VERSION AI413784.1 GI:4257288

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 398)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT On Apr 21, 1998 this sequence version replaced gi:3071892.

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end

High quality sequence stop: 375.

## FEATURES

Location/Qualifiers  
1..398  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:402963"  
/clone\_lib="Soares mouse embryo NbMEL3.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand CDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGCGGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]; double-stranded CDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo."

BASE COUNT 89 a 90 c 107 g 112 t

ORIGIN

## alignment\_scores:

Quality: 37.00 Length: 9  
Ratio: 4.625 Gaps: 0  
Percent Similarity: 88.889 Percent Identity: 88.889

## alignment\_block:

US-08-653-294-15 x AI413784

Align seg 1/1 to: AI413784 from: 1 to: 398

1 TyrArgLeuLeuIleArgLeuAspGlu 9

233 TACCTCTCTCTCATCAGACTGGATGAG 259

seq\_name: gb\_est15:AA465578

seq\_documentation\_block:

LOCUS AA465578 402 bp mRNA EST 15-AUG-1997  
 DEFINITION aa24f06.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:814211 5' similar to gb:X56253 rnal CATION-DEPENDENT MANNOSYL-6-PHOSPHATE RECEPTOR PRECURSOR (HUMAN); mRNA sequence.

ACCESSION AA465578

VERSION AA465578.1 GI:2191745

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 402)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On May 9, 1995 this sequence version replaced gi:802540.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28m3 rev2 Et from Amersham

High quality sequence stop: 313.

## FEATURES

source

1..402

/organism="Homo sapiens"

/db\_xref="GDB:6032206"

/db\_xref="taxon:9606"

/clone\_image="814211"

/clone\_lib="NCI\_CGAP\_GCB1"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD+),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

[5'-TGCTACCAATCTGAGTGGAGCGCGCTCATTTTCTTTTCTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 90 a 87 c 106 g 119 t

ORIGIN

## alignment\_scores:

Quality: 37.00 Length: 10

Ratio: 3.700 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 60.000

## alignment\_block:

US-08-653-294-15 x AA465578/rev ..

Align seg 1/1 to reverse of: AA465578 from: 1 to: 402

1 TyrArgLeuLeuIleArgLeuAspGluArg 10

|||||:|||||:|||||:|||||:|||||

205 TACCAGCTTTTACTACAATGGATGAAGA 176

seq\_name: gb\_est38:AW048963

seq\_documentation\_block:

LOCUS AW048963 407 bp mRNA EST 18-SEP-1999  
 DEFINITION UI-M-BH1-amo-g-04-0-UI-s1 NIH\_BMAP\_M\_S2 Mus musculus cDNA clone UI-M-BH1-amo-g-04-0-UI 3', mRNA sequence.

ACCESSION AW048963

VERSION AW048963.1 GI:5909492

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 407)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT On Mar 16, 1998 this sequence version replaced gi:2961735.

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestr@mail.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized pineal glands library cDNA library preparation: M.B.

Soares Lab Clone distribution: NIH BMAP cDNA clones will be made

available by the means that is soon to be determined. When NIH

determines the means for distribution of the BMAP cDNA clones, this

record will be updated accordingly when that means is determined.

Seq primer: M13 Forward

POLYA-res.

## FEATURES

source

1..407

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="NIH\_BMAP\_M\_S2"

/dev\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The

NIH\_BMAP\_M\_S2 library is a subtracted library derived from

NIH\_BMAP\_M\_S1, which in turn is a subtracted library

derived from a mixture of normalized libraries from ten

regions of the mouse brain (cerebellum, brain stems,

olfactory bulbs, hypothalamus, cortex, amygdala, basal

ganglia, pineal gland, striatum, hippocampus). The driver

used for subtraction consisted of a pool of 5,000 clones

from the NIH\_BMAP\_M\_S1 library and a pool of 2,000 clones

obtained from non-normalized and normalized mouse brain

spinal cord libraries.

TAG\_LIB=NIH\_BMAP\_M\_S2

TAG\_TISSUE=pineal-glands

TAG\_SEQ=AAGAC

BASE COUNT 91 a 91 c 105 g 120 t

ORIGIN

## alignment\_scores:

Quality: 37.00

Ratio: 4.625

Length: 9

Gaps: 0



Percent Similarity: 88.889 Percent Identity: 88.889

## alignment\_block:

US-08-653-294-15 x AW048963

Align seg 1/1 to: AW048963 from: 1 to: 407

1 TyrArgLeuLeuIleArgLeuAspGlu 9

|||||

239 TACCTCCCTCCATCAGACTGATGAG 265

seq\_name: gb\_est5:N28776

seq\_documentation\_block: 432 bp mRNA EST 04-JAN-1996  
LOCUS N28776  
DEFINITION YX69:09 r1 Soares melanocyte 2NBHM Homo sapiens cDNA clone  
IMAGE:267040 5' similar to gb:X56233\_rnal CATION-DEPENDENT  
MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (HUMAN); mRNA sequence.

ACCESSION N28776  
VERSION N28776.1 GI:1147012  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 432)

AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F., Trevasakis E., Waterston R., Williamson A., Wohlmann P. and Wilson R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Source: IMAGE Consortium, LLNL

High quality sequence stops: 323

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: T7

High quality sequence stop: 323.

Location/Qualifiers

1..432

/organism="Homo sapiens"

/db\_xref="GDB:3876682"

/db\_xref="taxon:9606"

/clone="IMAGE:267040"

/sex="Male"

/tissue\_type="melanocyte"

/lab\_host="DH10B (ampicillin resistant)"

/notes="Vector: pT73D (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAAGTGGAGCGCGCGAGTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaldo. RNA from normal foreskin melanocytes

(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 89 a 100 c 107 g 134 t 2 others

ORIGIN

alignment\_scores:

Quality: 37.00

Ratio: 3.700

Percent Similarity: 100.000 Percent Identity: 60.000

Length: 10

Gaps: 0

## alignment\_block:

US-08-653-294-15 x N28776/rev

Align seg 1/1 to reverse of: N28776 from: 1 to: 432

1 TyrArgLeuLeuIleArgLeuAspGlu 10

|||||

129 TACCAGCTTTTACTACAAATGATGATGAGA 100

seq\_name: gb\_est10:AA137143

seq\_documentation\_block: 433 bp mRNA EST 14-MAY-1997  
LOCUS AA137143  
DEFINITION 2102F08 r1 Soares pregnant uterus\_NbHPU Homo sapiens cDNA clone  
IMAGE:491175 5' similar to gb:X56253\_rnal CATION-DEPENDENT  
MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (HUMAN); mRNA sequence.

ACCESSION AA137143  
VERSION AA137143.1 GI:1698378  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 433)

AUTHORS Hillier L., Lennon G., Becker M., Bonaldo M.F., Chiapelli B.,

Chissoe S., Dietrich N., DuBuque T., Favello A., Gish W.,

Hawkins M., Hultman M., Kucaba T., Lacy M., Le N.,

Mardis E., Moore B., Morris M., Parsons J., Prange C., Rifkin L.,

Rohlfing T., Schellenberg K., Soares M.B., Tan F., Thierry-Mieg J.,

Trevasakis E., Underwood K., Wohlmann P., Waterston R., Wilson R.,

and Marra M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1393248.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 529 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 395.

Location/Qualifiers

1..433

/organism="Homo sapiens"

/db\_xref="GDB:3805762"

/db\_xref="taxon:9606"

/clone="IMAGE:491175"

/clone\_lib="Soares\_pregnant\_uterus\_NbHPU"

/sex="female"

/dev\_stage="adult"

/lab\_host="DH10B"

/notes="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5',

AACTGGAGAAATTCGCGCGCTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adapters

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

BASE COUNT 135 a 101 c 90 g 107 t

ORIGIN

alignment\_scores:

Quality: 37.00

Ratio: 3.700

Length: 10

Gaps: 0



```
alignment_scores:      Quality: 37.00      Length: 10
                       Ratio: 3.700        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 60.000

alignment_block:
US-08-653-294-15 x N46203/rev ..

Align seg 1/1 to reverse of: N46203 from: 1 to: 434

1 TyrArgLeuLeuIleArgLeuAspGluArg 10
||||:|||||:||||:||||:|||||
307 TACCAGCTTTTACTACAAATGGATGAAGA 278
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:29:38 ; Search time 122.56 Seconds  
(without alignments)  
2.319 Million cell updates/sec

Title: US-08-653-294-16  
Perfect score: 58  
Sequence: 1 YRLIRRLRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	48	82.8	12	1	R95429	HLA-B2702 84-79-84
2	48	82.8	12	1	W33798	Peptide B2702.84-7
3	48	82.8	12	1	W33799	Immunomodulating d
4	34	58.6	20	1	R32907	HLA-B2702 CTL modu
5	34	58.6	20	1	R95428	HLA-B2702 84-75-84
6	34	58.6	20	1	W33778	Immunomodulating d
7	34	58.6	361	1	R42281	PGE2 receptor codi
8	34	58.6	365	1	R42280	PGE2 receptor codi
9	34	58.6	365	1	R95518	Prostaglandin-EP3-
10	34	58.6	388	1	R69517	Prostaglandin-EP3-
11	34	58.6	390	1	R69516	Prostaglandin-EP3-
12	34	58.6	393	1	W57411	Human prostaglandi
13	34	58.6	402	1	W57410	Human EP3-v recept
14	33	56.9	3391	1	W06591	Polyprotein of att
15	33	56.9	3391	1	W06590	Polyprotein of DEN
16	32	55.2	263	1	P70467	Sequence of gpm en
17	32	55.2	644	1	W13821	Yeast transcriptio
18	32	55.2	660	1	W11725	H-Delta-1 polypept
19	32	55.2	2441	1	R79054	CREB binding prote
20	32	55.2	2441	1	W40058	Cellular transcrip
21	31	53.4	136	1	W41731	Actinobacillus ple
22	31	53.4	261	1	R21517	Polyfunctional pro
23	31	53.4	261	1	R22960	Human proteasome c
24	31	53.4	504	1	W44156	Human neuronal nic
25	31	53.4	599	1	R21690	Prostaglandin endo
26	31	53.4	904	1	W86351	Human DNAX toll-11
27	30	51.7	15	1	W1462	Conserved peptide
28	30	51.7	21	1	R30657	Peptide contg. alt
29	30	51.7	21	1	W88194	Surfactant peptide
30	30	51.7	40	1	W12037	Curvularia verrucu
31	30	51.7	180	1	R66185	Thai catfish GHRH/
32	30	51.7	216	1	R52029	Protein with Oxeta
33	30	51.7	221	1	W11703	OREF10 thermally st
34	30	51.7	221	1	W48738	OREF10 glycosylase.

35 30 51.7 221 1 W81465  
36 30 51.7 504 1 W09022  
37 30 51.7 600 1 W12042  
38 30 51.7 640 1 R82249  
39 30 51.7 761 1 W99790  
40 30 51.7 930 1 W55961  
41 30 51.7 4472 1 R97246  
42 29 50.0 6 1 W47264  
43 29 50.0 6 1 W47262  
44 29 50.0 6 1 W33782  
45 29 50.0 6 1 W33783

Thymine DNA glycos  
Neuronal nicotinic  
Curvularia verrucu  
Chloroperoxidase.  
Rat VRP-1 (VR2) c  
Human transient re  
Virulence gene clu  
Immunomodulatory p  
Immunomodulatory p  
Peptide #3 used in  
Peptide #4 used in

## ALIGNMENTS

### RESULT 1

R95429 standard; peptide; 12 AA.

AC R95429;  
DT 12-NOV-1996 (first entry)  
DE HLA-B2702 84-79-84 palindrome.  
KW HLA: p74; alpha-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN W09513288-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C. Krensky AM;  
DR WPI: 95-194027/25.  
PT Compsns. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12; 29pp; English.

CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. This sequence represents the  
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane  
CC protein associated with T-cell activation in mammalian T-cells, and is  
CC also immunologically cross reactive with the heat shock protein Hsc70.  
CC p74 is found in a limited number of cell types, but is particularly  
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
CC cell with an amphoteric detergent, and then passed through an affinity  
CC column containing a covalently bound HLA-B2702 palindromic peptide.  
CC Compositions comprising the extracellular fragment of p74 combined with  
CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
CC compounds can be screened for their effect on the cytolytic activity of  
CC T-cells, by combining them with the extracellular portion of p74 and  
CC determining the amount of binding between the candidate compound and p74.  
CC Modulation of CTL activity can be inhibited in a cellular composition  
CC containing T-cells and antigen presenting cells (APCs), by adding to the  
CC mix the extracellular portion of p74, in an amount sufficient to compete  
CC with p74 for the binding of the p74 ligand.  
SQ Sequence 12 AA;

Query Match 82.8%; Score 48; DB 1; Length 12;  
Best Local Similarity 83.3%; Pred. No. 0.0038;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIRRLRY 12  
||| |||| |||  
Db 1 YRLAIRRLRY 12

### RESULT 2

W33798 standard; peptide; 12 AA.

ID W33798  
AC W33798;

19-JUN-1998 (first entry)  
 DE Peptide B2702.84-79/79-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 82.8%; Score 48; DB 1; Length 12;  
 Best Local Similarity 83.3%; Pred. No. 0.0038;  
 Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIRRLRY 12  
 ||| ||| |||  
 DB 1 YRLAIRRLRY 12

RESULT 3  
 W33799  
 ID W33799 standard; peptide; 12 AA.  
 AC W33799;  
 DT 19-JUN-1998 (first entry)  
 DE Immunomodulating dimer peptide #3.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Claim 17; Page 35; 41pp; English.  
 CC This sequence represents a specifically claimed immunomodulating  
 CC dimer peptide of the invention. A peptide-type compound or variant is

CC claimed which has immunomodulating activity, including the N-terminal  
 CC acylated and/or C-terminal amidated or esterified forms of up to 60  
 CC amino acids, where the peptide-type compound comprises the formula: A-B,  
 CC where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or  
 CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
 CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa  
 CC represents amino acid. The sequence in the brackets may optionally be  
 CC absent or truncated at any peptide type bond within the brackets. The  
 CC compounds comprise amino acid sequences related to a Class I HLA-B  
 CC alpha domain (positions 79-84). They can be used to inhibit cytotoxic  
 CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in  
 CC vitro. They can also be used in combination with antigenic peptides or  
 CC proteins of interest to activate CTLs. They can also inhibit the  
 CC proliferation of T cells in response to anti-CD3. The peptide can be  
 CC used for preventing rejection of transplants or for treating autoimmune  
 CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.  
 CC The products can also be used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 82.8%; Score 48; DB 1; Length 12;  
 Best Local Similarity 83.3%; Pred. No. 0.0038;  
 Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIRRLRY 12  
 ||| ||| |||  
 DB 1 YRLAIRRLRY 12

## RESULT 4

R92907  
 ID R92907 standard; peptide; 20 AA.  
 AC R92907;  
 DT 16-MAY-1996 (first entry)  
 DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW class I MHC; HLA-B2702.  
 OS Synthetic.  
 PN WO9526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI: 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 15; Page 36; 80pp; English.  
 CC R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
 CC I MHC HLA-B2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with a  
 CC subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;  
 Best Local Similarity 50.0%; Pred. No. 2.4;  
 Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RILRY 12  
 ||| ||| ||| |||  
 DB 1 YRLAIRNERENLRALRY 20

```

RESULT 5
R95428 ID R95428 standard; peptide; 20 AA.
AC R95428:
DE 12-NOV-1996 (first entry)
KW HLA-B*2702 84-75-84 palindromic.
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW synthesis; antigen presenting cell.
OS Synthetic.
PN W09513288-A1.
PD 18-NOV-1995.
PF 10-NOV-1994; U12985.
PR 10-NOV-1993; US-150493.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI; 95-194027/25.
PT Compsns. comprising lymphoid surface membrane proteins - which may
PT inhibit cytolytic activity and differentiation of CTLs.
PS Example; Page 12; 29pp; English.
CC R95413, and R95415-R95431 represent palindromes and fragments of
CC human-leucocyte-associated antigens. This sequence represents the
CC HLA-B*2702 84-75-84 palindromic. These sequences can be used to isolate
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC protein associated with T-cell activation in mammalian T-cells, and is
CC also immunologically cross reactive with the heat shock protein Hsc70.
CC p74 is found in a limited number of cell types, but is particularly
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC cell with an amphoteric detergent, and then passed through an affinity
CC column containing a covalently bound HLA-B*2702 palindromic peptide.
CC Compositions comprising the extracellular fragment of p74 combined with
CC HLA-B*2702.60-84 (see R95416), induces calcium influx, and inhibits
CC cytotoxic T lymphocyte (CTL) differentiation or cytotoxicity. Candidate
CC compounds can be screened for their effect on the cytolytic activity of
CC T-cells, by combining them with the extracellular portion of p74 and
CC determining the amount of binding between the candidate compound and p74.
CC Modulation of CTL activity can be inhibited in a cellular composition
CC containing T-cells and antigen presenting cells (APCs), by adding to the
CC mix the extracellular portion of p74, in an amount sufficient to compete
CC with p74 for the binding of the p74 ligand.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIIR-----RILRY 12
   ||| ||
DB 1 YRLAIRLNRERLRY 20

RESULT 7
R42281 ID R42281 standard; Protein; 361 AA.
AC R42281;
DE 21-FEB-1994 (first entry)
DE PGE2 receptor coding sequence (Clone MP653).
KW Prostaglandin E receptor; prostaglandin; binding; agonist;
KW antagonist.
OS Mus musculus.
FH Key
FT domain 31..55 Location/Qualifiers
FT FT /label= Transmembrane domain 1.
FT FT 69..89
FT FT /label= Transmembrane domain 2.
FT FT 109..130
FT FT /label= Transmembrane domain 3.
FT FT 152..173
FT FT /label= Transmembrane domain 4.
FT FT 204..229
FT FT /label= Transmembrane domain 5.
FT FT 260..283
FT FT /label= Transmembrane domain 6.
FT FT 304..325
FT FT /label= Transmembrane domain 7.
PN EP-557966-A.
PD 01-SEP-1993.
PF 24-FEB-1993; 102873.
PR 24-FEB-1992; JP-036580.
PR 23-MAR-1992; JP-064889.
PI (TAKE ) TAKEDA CHEM IND LTD.
PI Ichikawa A, Narumiya S;
DR WPI; 93-274435/35.
DR N-PSDB; Q46125.
PT New prostaglandin E receptor protein and DNA encoding it - used
PT to study prostaglandin binding, agonists and antagonists etc.
PS Claim 5, 10; Figure 4; 22pp; English.
CC The prostaglandin E receptor can be used to clone other PGE
CC receptor genes, to clarify the structure of PGE receptors, to
CC elucidate the function of PGE and to identify PGE agonists/
CC antagonists. The receptor is produced by recombinant methods,

```

CC transforming a host cell, culturing the cell and retrieving the  
 CC recombinant protein from the culture medium. This PGE2 receptor is  
 CC a derivative of the clone MP660 (Q46124) having 89 base pairs  
 CC deleted from the coding sequence resulting in replacement of 30  
 CC C-terminal amino acids with a different sequence of 26 amino acids.  
 SQ Sequence 361 AA;

Query Match 58.6%; Score 34; DB 1; Length 361;  
 Best Local Similarity 77.8%; Pred. NO. 46;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11  
 ||:|:||||  
 Db 323 LLIRKILLR 331

RESULT 8  
 R42280 ID R42280 standard; Protein: 365 AA.  
 AC R42280;  
 DT 21-FEB-1994 (first entry)  
 DE PGE2 receptor coding sequence (Clone MP660).  
 KW Prostaglandin E receptor; prostaglandin; binding; agonist;  
 KW antagonist.  
 OS Mus musculus.  
 FH Key Location/Qualifiers  
 FT modified\_site 16  
 FT /label= N-glycosylation site.  
 FT domain 31..55  
 FT /label= Transmembrane domain 1.  
 FT modified\_site 59  
 FT /label= Phosphorylation site.  
 FT /note= Phosphorylation is by CAMP protein kinase.  
 FT modified\_site 64  
 FT /label= Phosphorylation site.  
 FT domain 69..89  
 FT /note= Phosphorylation is by CAMP protein kinase.  
 FT domain 109..130  
 FT /label= Transmembrane domain 2.  
 FT domain 152..173  
 FT /label= Transmembrane domain 3.  
 FT modified\_site 193  
 FT /label= Transmembrane domain 4.  
 FT domain 204..229  
 FT /label= N-glycosylation site.  
 FT domain 260..283  
 FT /label= Transmembrane domain 5.  
 FT domain 304..325  
 FT /label= Transmembrane domain 6.  
 FT domain 304..325  
 FT /label= Transmembrane domain 7.  
 PN EP-557966-A.  
 PD 01-SEP-1993.  
 PF 24-FEB-1993. 102873.  
 PR 24-FEB-1992; JP-036580.  
 PR 23-MAR-1992; JP-064889.  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 PI Ichikawa A, Narumiya S;  
 DR WPI: 93-274435/35.  
 DR N-PSDB: Q46124.  
 PT New prostaglandin E receptor protein and DNA encoding it - used  
 PT to study prostaglandin binding, agonists and antagonists etc.  
 PS Claim 10; Figure 1; 22pp; English.  
 CC The prostaglandin E receptor can be used to clone other PGE  
 CC receptor genes, to clarify the structure of PGE receptors, to  
 CC elucidate the function of PGE and to identify PGE agonists/  
 CC antagonists. The receptor is produced by recombinant methods,  
 CC transforming a host cell, culturing the cell and retrieving the  
 CC recombinant protein from the culture medium.  
 SQ Sequence 365 AA;

Query Match 58.6%; Score 34; DB 1; Length 365;

Best Local Similarity 77.8%; Pred. NO. 46;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 LLIRILLR 11  
 ||:|:||||  
 Db 323 LLIRKILLR 331

RESULT 9  
 R69518 ID R69518 standard; Protein: 365 AA.  
 AC R69518;  
 DT 02-SEP-1995 (first entry)  
 DE Prostaglandin-EP3-9 receptor.  
 KW Human uterus prostaglandin-EP3-9 receptor;  
 KW prostaglandin-EP3 receptor-agonist;  
 KW prostaglandin-EP3 receptor-antagonist.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT modified\_site 18  
 FT /note= "N-glycosylation site"  
 FT modified\_site 36  
 FT /note= "N-glycosylation site"  
 FT modified\_site 217  
 FT /note= "N-glycosylation site"  
 FT modified\_site 308  
 FT /note= "N-glycosylation site"  
 FT region 338..345  
 FT /note= "conserved sequence in EP3 receptors"  
 PN W09500552-A.  
 PD 05-JAN-1995.  
 PF 09-JUN-1994; CA0320.  
 PR 25-JUN-1993; US-083746.  
 PA (MERL ) MERCK FROSST CANADA INC.  
 PI Abramovitz M, Adam M, Boie Y, Metters K, Rushmore TH;  
 DR WPI: 93-052012/07.  
 DR N-PSDB: Q81951.  
 PT New prostaglandin EP3 receptors and DNA - used partic. to  
 PT identify modulators of prostaglandin receptor activity for  
 PT treatment of diseases.  
 PS Claim 2; Page 43-45; 64pp; English.  
 CC The sequence represents a human uterus prostaglandin-EP3-9  
 CC receptor (mol.wt. 40,507). The sequence contains 4 conserved  
 CC N-glycosylation sites in putative extracellular regions, and a  
 CC conserved region (R69515) in transmembrane region-VII common to  
 CC other EP3 receptors. Conserved Cys residues are found in  
 CC exofacial loops 1 and 2, and a conserved Arg found in all  
 CC eicosanoid receptors is found in transmembrane region-VII. The  
 CC receptor may be used to identify receptor-agonists and  
 CC receptor-antagonists, for use in therapy of e.g. glaucoma,  
 CC side-effects of non-steroidal antiinflammatories, cancer,  
 CC metastasis, renal vasoconstriction, etc.  
 SQ Sequence 365 AA;

Query Match 58.6%; Score 34; DB 1; Length 365;  
 Best Local Similarity 77.8%; Pred. NO. 46;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11  
 ||:|:||||  
 Db 347 LLIRKILLR 355

RESULT 10  
 R69517 ID R69517 standard; Protein: 388 AA.  
 AC R69517;  
 DT 02-SEP-1995 (first entry)  
 DE Prostaglandin-EP3-21 receptor.  
 KW Human uterus prostaglandin-EP3-21 receptor;  
 KW prostaglandin-EP3 receptor-agonist;  
 KW prostaglandin-EP3 receptor-antagonist.



```

OS Homo sapiens.
FH Key Location/Qualifiers
FT modified_site 18
FT /note= "N-glycosylation site"
FT modified_site 36
FT /note= "N-glycosylation site"
FT modified_site 217
FT /note= "N-glycosylation site"
FT modified_site 308
FT /note= "N-glycosylation site"
FT region 338..345
FT /note= "conserved sequence in EP3 receptors"
FT WO9500552-A.
PN PD 05-JAN-1995.
PF 09-JUN-1994; CA0320.
PR 25-JUN-1993; US-083746.
PA (MERI ) MERCK FROST CANADA INC.
PI Abramovitz M, Adam M, Boie Y, Metters K, Rushmore TH;
DR WPI: 95-052012/07.
DR N-PSDB; Q81949.
PT New prostaglandin EP3 receptors and DNA - used partic. to
PT identify modulators of prostaglandin receptor activity for
PT treatment of diseases.
PS Claim 2; Page 42-43; 64pp; English.
CC The sequence represents a human kidney prostaglandin-EP3-alpha
CC receptor (mol.wt. 42,688). The sequence contains 4 conserved
CC N-glycosylation sites in putative extracellular regions, and a
CC conserved region (R69515) in transmembrane region-VII which has
CC been used to generate a probe for isolation of other EP3 receptor
CC genes. Conserved Cys residues are found in exofacial loops 1 and
CC 2, and a conserved Arg found in all eicosanoid receptors is found
CC in transmembrane region-VII. The receptor may be used to identify
CC receptor-agonists and receptor-antagonists, for use in therapy of
CC e.g. glaucoma, side-effects of non-steroidal antiinflammatories,
CC cancer, metastasis, renal vasoconstriction, etc.
SQ Sequence 388 AA;

Query Match 58.6%; Score 34; DB 1; Length 388;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11
DB 347 LLRKILLR 355
|||||

RESULT 11
ID R69516 standard; Protein; 390 AA.
AC R69516;
DT 02-SEP-1995 (first entry)
DE Prostaglandin-EP3-alpha receptor.
KW Human kidney prostaglandin-EP3-alpha receptor;
KW prostaglandin-EP3 receptor-agonist;
KW prostaglandin-EP3 receptor-antagonist.
OS Homo sapiens.
PN J1013185-A.
PD 06-MAY-1998.
PF 14-OCT-1996; 291150.
PR 14-OCT-1996; JP-291150.
PA (ONOV ) ONO PHARM CO LTD.
DR WPI: 98-315474/28.
DR N-PSDB; V29611.
PT New human prostaglandin EP3 receptor(s) - useful for treatment and
PT prevention of, e.g. inflammation
PS Claim 8; Pages 20-21; 27pp; Japanese.
CC This represents a human EP3-VI receptor. A replication or expression
CC vector comprising cDNA sequences encoding EP-3V or EP3-3VI can be used
CC to transform a host cell. The host cell is cultured and the polypeptides
CC can be recovered from the culture medium. The polypeptides combine
CC specifically with a prostaglandin PGE2 receptor and can be used as a
CC preventive and treating agent for inflammation.
SQ Sequence 393 AA;

Query Match 58.6%; Score 34; DB 1; Length 393;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11
DB 347 LLRKILLR 355
|||||

RESULT 13
ID W57410 standard; Protein; 402 AA.
AC W57410;
DT 19-AUG-1998 (first entry)

```

```

PA (MERI ) MERCK FROST CANADA INC.
PI Abramovitz M, Adam M, Boie Y, Metters K, Rushmore TH;
DR WPI: 95-052012/07.
DR N-PSDB; Q81949.
PT New prostaglandin EP3 receptors and DNA - used partic. to
PT identify modulators of prostaglandin receptor activity for
PT treatment of diseases.
PS Claim 2; Page 40-41; 64pp; English.
CC The sequence represents a human kidney prostaglandin-EP3-alpha
CC receptor (mol.wt. 43,315). The sequence contains 4 conserved
CC N-glycosylation sites in putative extracellular regions, and a
CC conserved region (R69515) in transmembrane region-VII which has
CC been used to generate a probe for isolation of other EP3 receptor
CC genes. Conserved Cys residues are found in exofacial loops 1 and
CC 2, and a conserved Arg found in all eicosanoid receptors is found
CC in transmembrane region-VII. The receptor may be used to identify
CC receptor-agonists and receptor-antagonists, for use in therapy of
CC e.g. glaucoma, side-effects of non-steroidal antiinflammatories,
CC cancer, metastasis, renal vasoconstriction, etc.
SQ Sequence 390 AA;

Query Match 58.6%; Score 34; DB 1; Length 390;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11
DB 347 LLRKILLR 355
|||||

RESULT 12
ID W57411 standard; Protein; 393 AA.
AC W57411;
DT 19-AUG-1998 (first entry)
DE Human prostaglandin EP3-VI receptor.
KW Prostaglandin E2 receptor; EP3-V receptor; human; treatment;
KW inflammation; EP3-VI.
OS Homo sapiens.
PN J1013185-A.
PD 06-MAY-1998.
PF 14-OCT-1996; 291150.
PR 14-OCT-1996; JP-291150.
PA (ONOV ) ONO PHARM CO LTD.
DR WPI: 98-315474/28.
DR N-PSDB; V29611.
PT New human prostaglandin EP3 receptor(s) - useful for treatment and
PT prevention of, e.g. inflammation
PS Claim 8; Pages 20-21; 27pp; Japanese.
CC This represents a human EP3-VI receptor. A replication or expression
CC vector comprising cDNA sequences encoding EP-3V or EP3-3VI can be used
CC to transform a host cell. The host cell is cultured and the polypeptides
CC can be recovered from the culture medium. The polypeptides combine
CC specifically with a prostaglandin PGE2 receptor and can be used as a
CC preventive and treating agent for inflammation.
SQ Sequence 393 AA;

Query Match 58.6%; Score 34; DB 1; Length 393;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11
DB 347 LLRKILLR 355
|||||

RESULT 13
ID W57410 standard; Protein; 402 AA.
AC W57410;
DT 19-AUG-1998 (first entry)

```

```

DE Human EP3-V receptor.
KW Prostaglandin E2 receptor; EP3-V receptor; human; treatment;
KW inflammation; EP3-VI.
OS Homo sapiens.
PN J10113185-A.
PD 06-MAY-1998.
PF 14-OCT-1996; 2911150.
PR 14-OCT-1996; JP-2911150.
PA (ONONY ) ONO PHARM CO LTD.
DR WPI: 98-315474/28.
DR N-PSDB: V29610.
PT New human prostaglandin EP3 receptor(s) - useful for treatment and
PT prevention of, e.g. inflammation.
PS Claim 2; pages 16-17; 27pp; Japanese.
CC This represents a human EP3-V receptor. A replication or expression
CC vector comprising cDNA sequences encoding EP-3V or EP3-3VI can be used
CC to transform a host cell. The host cell is cultured and the polypeptides
CC can be recovered from the culture medium. The polypeptides combine
CC specifically with a prostaglandin PGE2 receptor and can be used as a
CC preventive and treating agent for inflammation.
SQ Sequence 402 AA;

Query Match 58.6%; Score 34; DB 1; Length 402;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11
DB 347 LLLRILLR 355
||:|||||

RESULT 14
W06591
ID W06591 standard; Protein; 3391 AA.
AC W06591;
DE 12-SEP-1997 (first entry)
DE Polypeptide of attenuated DEN-2 virus, strain 16681, PDK-53.
DE Dengue 2 virus; polypeptide; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;
KW NS4A; NS4B; NS5; PDK-53; quadravalent vaccine; immunity; serotype;
KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;
KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;
KW DHF; DSS.
OS Dengue 2 virus, strain 16681.
FH Key Location/Qualifiers
FT protein 2..114
FT /label= Capsid_protein
FT 115..205
FT /label= prM
FT modified_site 183
FT /note= "N-linked glycosylation site, encoded by NAC"
FT protein 206..280
FT /label= M
FT protein 281..775
FT /label= E
FT modified_site 347
FT /note= "N-linked glycosylation site, encoded by NAC"
FT modified_site 433
FT /note= "N-linked glycosylation site, encoded by NAT"
FT protein 776..1127
FT /label= NS1
FT modified_site 905
FT /note= "N-linked glycosylation site, encoded by NAC"
FT modified_site 982
FT /note= "N-linked glycosylation site, encoded by NAT"
FT protein 1128..1345
FT /label= NS2A
FT protein 1346..1475
FT /label= NS2B
FT protein 1476..2093
FT /label= NS3
FT protein 2094..2242
FT /label= NS4A

```

```

FT protein 2243..2491
FT /label= NS4B
FT 2492..3391
FT /label= NS5
FT misc_difference 3038
FT /note= "Encoded by KKA"
PN W09640933-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09209.
PR 07-JUN-1995; US-483292.
PA (UYMA-) UNIV MAHIDOL AT SALAYA.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Bhamarapratvi N, Butrapet S, Chang J, Gubler DJ;
PI Halstead SB, Kinney R, Trent DW;
DR WPI: 97-052330/05.
DR N-PSDB: T49304.
PT PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681
PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a
PT quadravalent vaccine for protecting against Dengue virus infection
PS Claim 27; page 122-136; 261pp; English.
CC This sequence represents the polypeptide from attenuated Dengue 2 virus,
CC strain 16681. The attenuated virus is designated PDK-53. The poly-
CC protein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3, NS4A,
CC NS4B and NS5 proteins. A clone of this wildtype viral sequence, PDK-53,
CC may be used in the production of a quadravalent vaccine which provides
CC immunity against all four serotypes of dengue virus. The vaccine also
CC comprises a chimeric DEN-2/1 virus, a chimeric DEN-2/3 virus, and/or a
CC chimeric DEN-2/4 virus. The new quadravalent vaccines are used to
CC protect against infection by all four serotypes of dengue virus, DEN-1,
CC DEN-2, DEN-3 and DEN-4, which can lead to dengue fever or fatal dengue
CC haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host cells are
CC used to produce the recombinant protein products of the DNA constructs
CC which are used in the vaccines.
SQ Sequence 3391 AA;

Query Match 56.9%; Score 33; DB 1; Length 3391;
Best Local Similarity 54.5%; Pred. No. 6.8e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLLIRILLRY 12
DB 2728 RMLIRFTMYR 2738
||:|||||

RESULT 15
W06590
ID W06590 standard; Protein; 3391 AA.
AC W06590;
DE 11-SEP-1997 (first entry)
DE Polypeptide of DEN-2 virus, strain 16681.
DE Dengue 2 virus; polypeptide; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;
KW NS4A; NS4B; NS5; PDK-53; quadravalent vaccine; immunity; serotype;
KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;
KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;
KW DHF; DSS.
OS Dengue 2 virus, strain 16681.
FH Key Location/Qualifiers
FT protein 2..114
FT /label= Capsid_protein
FT 115..205
FT /label= prM
FT modified_site 183
FT /note= "N-linked glycosylation site"
FT protein 206..280
FT /label= M
FT protein 281..775
FT /label= E
FT modified_site 347
FT /note= "N-linked glycosylation site"
FT modified_site 433
FT /note= "N-linked glycosylation site"
FT protein 776..1127

```

```

FT modified_site /label= NS1
FT 905
FT /note= "N-linked glycosylation site"
FT modified_site 982
FT /note= "N-linked glycosylation site"
FT protein 1128..1345
FT /label= NS2A
FT protein 1346..1475
FT /label= NS2B
FT protein 1476..2093
FT /label= NS3
FT protein 2094..2242
FT /label= NS4A
FT protein 2243..2491
FT /label= NS4B
FT protein 2492..3391
FT /label= NS5
FT misc_difference 3038
FT /note= "Encoded by KKA"
FT WO9640933-Al.
FT 19-DEC-1996.
FT PD 06-JUN-1996; U09209.
FT PF 07-JUN-1995; US-483292.
FT PR (UYMA-) UNIV MAHIDOL AT SALAYA.
FT PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
FT PI Bhamarapavati N, Butrapet S, Chang J, Gubler DJ;
FT PI Halstead SB, Kinney R, Trent DW;
FT DR WPI: 97-052330/05.
FT DR N-PSDB: T49303.
FT PT PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681
FT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a
FT quadravalent vaccine for protecting against Dengue virus infection
FT Claim 23; Page 107-121; 261pp; English.
FT CC This sequence represents the polyprotein from Dengue 2 virus, strain
CC NS2B, NS3, NS4A, NS4B and NS5 proteins. A clone of this wildtype
CC viral sequence, PDK-53, may be used in the production of a quadravalent
CC vaccine which provides immunity against all four serotypes of dengue
CC virus. The vaccine also comprises a chimeric DEN-2/1 virus, a chimeric
CC DEN-2/3 virus, and/or a chimeric DEN-2/4 virus. The new quadravalent
CC vaccines are used to protect against infection by all four serotypes of
CC dengue virus, DEN-1, DEN-2, DEN-3 and DEN-4, which can lead to dengue
CC fever or fatal dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS).
CC Host cells are used to produce the recombinant protein products of the
CC DNA constructs which are used in the vaccines.
SQ Sequence 3391 AA;

```

```

Query Match 56.9%; Score 33; DB 1; Length 3391;
Best Local Similarity 54.5%; Pred. No. 6.8e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 RLLIRILLRY 12
Db 2728 RMLINRFTRY 2738

```

```

Search completed: February 8, 2000, 01:29:39
Job time: 1751 sec

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 11:54:26 ; Search time 117.7 seconds  
(without alignments)  
4.809 Million cell updates/sec

Title: US-08-653-294-16  
Perfect score: 58  
Sequence: 1 YRLIRILLR 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR62:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	63.8	408	2 D70365	fimbrial assembly
2	36	62.1	348	2 S29990	histocompatibility
3	36	62.1	559	2 E71490	probable efflux pr
4	36	62.1	1055	2 F64114	type I site-specif
5	35.5	61.2	278	2 S48776	hypothetical prote
6	35	60.3	298	2 S53849	ribosomal protein
7	35	60.3	814	2 T00953	hypothetical prote
8	35	60.3	1712	2 C71618	hypothetical prote
9	34	58.6	39	2 C55995	prostaglandin E2 r
10	34	58.6	48	2 D55995	prostaglandin E2 r
11	34	58.6	62	2 B55995	prostaglandin E2 r
12	34	58.6	78	2 F64429	DNA-directed RNA p
13	34	58.6	78	2 C53216	prostaglandin E2 r
14	34	58.6	185	1 RVN222	DNA-directed RNA p
15	34	58.6	185	2 F36845	DNA-directed RNA p
16	34	58.6	185	2 J46667	TB2/DP1 protein ho
17	34	58.6	213	2 D71669	adenylate kinase (
18	34	58.6	217	2 D53216	prostaglandin E2 r
19	34	58.6	313	2 JC5342	Na+/H+ antiporter
20	34	58.6	352	2 S38147	hypothetical prote
21	34	58.6	361	2 A45211	prostaglandin E2 r
22	34	58.6	361	2 A53216	prostaglandin E2 r
23	34	58.6	362	2 S48699	prostaglandin E(2)
24	34	58.6	364	2 S65009	prostaglandin E2 r
25	34	58.6	364	2 JC2115	prostaglandin E2 r
26	34	58.6	365	2 J38718	prostaglandin recep
27	34	58.6	365	2 A42414	prostaglandin E2 r
28	34	58.6	365	2 JN0693	prostaglandin E2 r
29	34	58.6	365	2 S51315	prostaglandin E2 r
30	34	58.6	366	2 S51280	EP3-alpha receptor

31	34	58.6	367	2 JC2056	prostaglandin E2 r
32	34	58.6	374	2 I38747	prostaglandin recep
33	34	58.6	374	2 S51317	prostaglandin E2 r
34	34	58.6	388	2 I38750	prostaglandin recep
35	34	58.6	388	2 S51316	prostaglandin E2 r
36	34	58.6	390	2 S43375	prostaglandin E2 r
37	34	58.6	390	2 S51313	prostaglandin E2 r
38	34	58.6	393	2 S51318	prostaglandin E2 r
39	34	58.6	411	2 B53216	prostaglandin E2 r
40	34	58.6	425	2 S51319	prostaglandin E2 r
41	34	58.6	575	2 B64174	hypothetical prote
42	34	58.6	735	2 T13646	hypothetical prote
43	34	58.6	928	2 S50578	hypothetical prote
44	34	58.6	3434	1 GNVNVV	genome polyprotein
45	33	56.9	133	2 D72110	hypothetical prote

## ALIGNMENTS

## RESULT 1

D70365  
fimbrial assembly protein PilC - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 20-Aug-1999  
C:Accession: D70365  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.

Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: D70365  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-408 <AQF>  
A:Cross-references: GB:AF000705; NID:g2983310; PIDN:AA06902.1; PID:g2983312; GB:AE00  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: pilC1  
C:Superfamily: secretion protein xcps

Query Match 63.8%; Score 37; DB 2; Length 408;  
Best Local Similarity 72.7%; Pred. No. 14;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YRLIRILLR 11  
: | | | | | | | | | |  
Db 250 FRALIRILLR 260

## RESULT 2

S29990  
histocompatibility antigen, HLA-F-like - rhesus macaque  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S29990  
R:Bontrop, R.R.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S29990  
A:Accession: S29990  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-348 <BON>  
A:Cross-references: EMBL:Z21819; NID:g38568; PIDN:CAA79885.1; PID:g38569  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:219-284/Domain: immunoglobulin homology <IMM>

Query Match 62.1%; Score 36; DB 2; Length 348;  
Best Local Similarity 54.5%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLLIRILLRY 12  
 I: : : : :  
 Db 98 RVALRKLLRY 108  
 I: : : : :  
 RESULT 3  
 E71490  
 Probable efflux protein - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
 C:Species: Chlamydia trachomatis  
 C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Nov-1998  
 C:Accession: E71490  
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998  
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
 A:Reference number: A71570; MUID:99000809  
 A:Accession: E71490  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-559 <ARN>  
 A:CROSS-references: GB:AE001334; GB:AE001273; MID:g3329078; PID:g3329089  
 A:Experimental source: serotype D, strain UW-3/Cx  
 C:Genetics:  
 A:Gene: ygeD  
 Query Match 62.1%; Score 36; DB 2; Length 559;  
 Best Local Similarity 63.8%; Pred. No. 30;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 YRLIRILLR 11  
 I: : : : :  
 Db 409 YRLIRVLR 419  
 I: : : : :  
 RESULT 4  
 F64114  
 Type I site-specific deoxyribonuclease (EC 3.1.21.3) chain hsdR - Haemophilus influenzae  
 C:Species: Haemophilus influenzae  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
 C:Accession: F64114  
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995  
 A:Authors: Guehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630  
 A:Accession: F64114  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1055 <TIGR>  
 A:CROSS-references: GB:U32808; GB:L42023; MID:g1574739; PID:g1574743; TIGR:H11285  
 C:Genetics:  
 A:Gene: hsdR  
 C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology  
 C:Keywords: ATP; DNA binding; hydrolase; P-loop; restriction modification system  
 F:300-665/Domain: DEAD/H box helicase homology <DEAD>  
 F:404-409/Region: nucleotide-binding motif B  
 F:408-411/Region: DEAD motif  
 Query Match 62.1%; Score 36; DB 2; Length 1055;  
 Best Local Similarity 63.8%; Pred. No. 56;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 RLLIRILLRY 12  
 I: : : : :  
 Db 1017 RLLVRRALQRY 1027  
 I: : : : :  
 RESULT 5  
 S48776  
 hypothetical protein YDR087c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein D4478; hypothetical protein YD8554.20c  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 07-May-1999  
 C:Accession: S48776; S49842; S39583; S55836; S67904  
 R:Coster, F.; Jonniaux, J.L.; Goffeau, A.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48758  
 A:Accession: S48776  
 A:Molecule type: DNA  
 A:Residues: 1-278 <COS>  
 A:CROSS-references: EMBL:X82086; NID:g558241; PID:g558260  
 R:Richards, C.; Harris, D.E.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S49823  
 A:Accession: S49842  
 A:Molecule type: DNA  
 A:Residues: 1-278 <RIC>  
 R:Esnault, Y.; Blondel, M.O.; Deshaies, R.J.; Scheekman, R.; Kepes, F. EMBO J. 12, 4083-4093, 1993  
 A:Title: The yeast SSI1 gene is essential for secretory protein translocation and enc  
 A:Reference number: S39583; MUID:94038890  
 A:Accession: S39583  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 146-278 <ESN>  
 A:CROSS-references: EMBL:X74499; NID:g414690; PID:g414691  
 R:Coster, F.; Jonniaux, J.L.; Goffeau, A.  
 yeast 11, 673-679, 1995  
 A:Title: Analysis of a 32.8 kb segment of yeast chromosome IV reveals 21 open reading  
 A:Reference number: S55819; MUID:96093910  
 A:Accession: S55836  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-278 <COW>  
 A:CROSS-references: EMBL:X82086; NID:g558241; PID:g558260  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994  
 R:Foury, F.; Jonniaux, J.L.; Purnelle, B.; Coster, F.; Goffeau, A.  
 submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67889  
 A:Accession: S67904  
 A:Molecule type: DNA  
 A:Residues: 1-278 <FOU>  
 A:CROSS-references: EMBL:274383; NID:g1431562; PID:e253404; PID:g1431563; MIPS:YDR087  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Map position: 4R  
 Query Match 61.2%; Score 35.5; DB 2; Length 278;  
 Best Local Similarity 60.0%; Pred. No. 19;  
 Matches 9; Conservative 2; Mismatches 1; Indels 3; Gaps 1;  
 QY 1 YRLIRILLR---LRY 12  
 I: : : : :  
 Db 121 YLLIRRVLFSLQKY 135  
 I: : : : :  
 RESULT 6  
 S53849  
 ribosomal protein S3 - Acanthamoeba castellanii mitochondrion (SGC6)  
 C:Species: mitochondrion Acanthamoeba castellanii  
 C>Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 26-Feb-1999  
 C:Accession: S53849  
 R:Burger, G.; Pianté, I.; Loneragan, K.M.; Gray, M.W. J. Mol. Biol. 245, 522-537, 1995  
 A:Title: The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba castellanii: c  
 A:Reference number: S53825; MUID:95147275  
 A:Accession: S53849  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-298 <BUR>

A:Cross-references: GB:U12386; NID:9562028; PID:9562053  
A:Experimental source: strain Neff; ATCC 30010  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC6  
C:Keywords: mitochondrion

Query Match 60.3%; Score 35; DB 2; Length 298;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIRILLRY 12  
: ||: ||: |||  
DB 126 FMLKKRVILRY 137

## RESULT 7

hypothetical protein F20D22.3 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Sep-1999

C:Accession: T00953

R:Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji, C.  
K.; Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.;  
submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence.

A:Reference number: Z14214

A:Accession: T00953

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-814 <YVS>

A:Cross-references: EMBL:AC002411; NID:g2570223; PID:g3142290; GSPDB:GN0059; ATSP:F20D22

C:Genetics:

A:Gene: ATSP:F20D22.3

A:Map position: 1

A:Introns: 64/1; 144/1; 239/3; 304/1; 386/1; 415/2; 473/3; 516/3; 554/3; 594/2; 624/3; 6

Query Match 60.3%; Score 35; DB 2; Length 814;  
Best Local Similarity 60.0%; Pred. No. 67;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLIRILLRY 12  
: ||: ||: |||  
DB 263 LVVRLILLY 272

## RESULT 8

hypothetical protein PFB0315w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 07-May-1999

C:Accession: C71618

R:Gardner, M.J.; Tetteilin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Perteira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;  
Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743

A:Accession: C71618

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1712 <GAP>

A:Cross-references: GB:AE001386; GB:AE001362; NID:g3845148; PID:g3845151; TIGR:PFB0315w

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0315w

Query Match 60.3%; Score 35; DB 2; Length 1712;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRILLRY 12  
: ||: ||: |||  
DB 1316 YKIYLRILFGY 1327

## RESULT 9

C55995

prostaglandin E2 receptor, subtype EP3 splice form III - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 17-Mar-1999

C:Accession: C55995

R:An, S.; Yang, J.; So, S.W.; Zeng, L.; Goetzl, E.J.

Biochemistry 33, 14496-14502, 1994

A:Title: Isoforms of the EP3 subtype of human prostaglandin E-2 receptor transduce bo

A:Reference number: A55995; MUID:95072021

A:Accession: C55995

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-39 <ANA>

A:Cross-references: GB:L32661

C:Superfamily: prostaglandin E receptor EP1

C:Keywords: alternative splicing

Query Match 58.6%; Score 34; DB 2; Length 39;  
Best Local Similarity 77.8%; Pred. No. 5.2;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11  
: ||: ||: |||  
DB 22 LLLRKILLR 30

## RESULT 10

D55995

prostaglandin E2 receptor, subtype EP3 splice form IV - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 17-Mar-1999

C:Accession: D55995

R:An, S.; Yang, J.; So, S.W.; Zeng, L.; Goetzl, E.J.

Biochemistry 33, 14496-14502, 1994

A:Title: Isoforms of the EP3 subtype of human prostaglandin E-2 receptor transduce bo

A:Reference number: A55995; MUID:95072021

A:Accession: D55995

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-48 <ANA>

A:Cross-references: GB:L32662

C:Superfamily: prostaglandin E receptor EP1

C:Keywords: alternative splicing

Query Match 58.6%; Score 34; DB 2; Length 48;  
Best Local Similarity 77.8%; Pred. No. 6.4;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11  
: ||: ||: |||  
DB 22 LLLRKILLR 30

## RESULT 11

B55995

prostaglandin E2 receptor, subtype EP3 splice form II - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 17-Mar-1999

C:Accession: B55995

R:An, S.; Yang, J.; So, S.W.; Zeng, L.; Goetzl, E.J.

Biochemistry 33, 14496-14502, 1994

A:Title: Isoforms of the EP3 subtype of human prostaglandin E-2 receptor transduce bo

A:Reference number: A55995; MUID:95072021

A:Accession: B55995

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-62 <ANA>  
A:Cross-references: GB:L32660  
C:Superfamily: prostaglandin E receptor EP1  
C:Keywords: alternative splicing

Query Match 58.6%; Score 34; DB 2; Length 62;  
Best Local Similarity 77.8%; Pred. No. 8.2;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LLIRILLR 11  
||:|:||||  
Db 22 LLLRKILLR 30

## RESULT 12

F64429  
DNA-directed RNA polymerase (EC 2.7.7.6) subunit H - Methanococcus jannaschii

C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 18-Jun-1999  
C:Accession: F64429  
R:Buit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999

A:Accession: F64429

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-78 <BUL>

A:Cross-references: GB:U67546; GB:L77117; NID:g1591687; PIDN:AAB99042.1; PID:g1591692; T

C:Genetics:

A:Map position: FOR970764-971000

A:Start codon: TTG

C:Superfamily: DNA-directed RNA polymerase chain H

C:Keywords: nucleotidyltransferase; transcription

Query Match 58.6%; Score 34; DB 2; Length 78;  
Best Local Similarity 66.7%; Pred. No. 10;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YRLIRILL 9  
||:|:||||  
Db 70 YRLVIKRII 78

## RESULT 13

C53216

prostaglandin E2 receptor, subtype EP3 (clone 74A) - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 23-Jul-1999

C:Accession: C53216

R:Breyer, R.M.; Emeson, R.B.; Tarnag, J.L.; Breyer, M.D.; Davis, L.S.; Abramson, R.M.; Re

J. Biol. Chem. 269, 6163-6169, 1994

A:Title: Alternative splicing generates multiple isoforms of a rabbit prostaglandin E-2

A:Reference number: A53216; MUID:94164982

A:Accession: C53216

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-78 <BRE>

A:Cross-references: GB:U04275; NID:g467982; PIDN:AA17414.1; PID:g467983

C:Superfamily: prostaglandin E receptor EP1

C:Keywords: alternative splicing; transmembrane protein

Query Match 58.6%; Score 34; DB 2; Length 78;  
Best Local Similarity 77.8%; Pred. No. 10;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LLIRILLR 11  
||:|:||||  
Db 44 LLLRKILLR 52

## RESULT 14

RNV222

DNA-directed RNA polymerase (EC 2.7.7.6) 22K chain - vaccinia virus (strains WR and C

N;Alternate names: J4R protein

C:Species: vaccinia virus

A:Note: host Homo sapiens (man)

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 11-Jun-1999

C:Accession: A25734; I23092; G42513

R:Broyles, S.S.; Moss, B.

Proc. Natl. Acad. Sci. U.S.A. 83, 3141-3145, 1986

A:Title: Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes:

A:Reference number: A25734; MUID:86205852

A:Accession: A25734

A:Molecule type: DNA

A:Residues: 1-185 <BRO>

A:Cross-references: GB:M13209; GB:M14122; NID:g335739; PIDN:AAB59833.1; PID:g335740

A:Experimental source: strain WR

R:Plucieniczak, A.; Schroeder, E.; Zettlmeissl, G.; Strecek, R.E.

Nucleic Acids Res. 13, 985-998, 1985

A:Title: Nucleotide sequence of a cluster of early and late genes in a conserved segm

A:Reference number: A23092; MUID:85215527

A:Accession: I23092

A:Molecule type: DNA

A:Residues: 1-185 <PLU>

A:Cross-references: GB:X01978; GB:J02424; GB:J02425; GB:K02376; GB:M15211; GB:V01537;

A:Experimental source: strain WR

R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paolletti, E.

Virology 179, 517-563, 1990

A:Title: Appendix to "The complete DNA sequence of vaccinia virus".

A:Reference number: A42501

A:Accession: G42513

A:Molecule type: DNA

A:Residues: 1-185 <GOE>

A:Cross-references: GB:M35027; NID:g335317; PIDN:AAA48084.1; PID:g335432

A:Experimental source: strain Copenhagen

R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paolletti, E.

Virology 179, 247-286, 1990

A:Title: The complete DNA sequence of vaccinia virus.

A:Reference number: A42531; MUID:91021027

A:Contents: annotation; possible protein-coding frames

A:Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: vaccinia virus DNA-directed RNA polymerase 22K polypeptide

C:Keywords: early protein; nucleotidyltransferase; transcription

Query Match 58.6%; Score 34; DB 1; Length 185;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YRLIRILLRY 12  
||:|:||||  
Db 24 YAVINRVLLRY 35

## RESULT 15

F36845

DNA-directed RNA polymerase (EC 2.7.7.6) chain J4R - variola virus

N:Alternate names: L4R protein

C:Species: variola virus

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 18-Jun-1999

C:Accession: F36845; S33095

R:Blinov, V.M.

submitted to GenBank, November 1992

A:Description: not shown.

A:Reference number: A36859

A:Accession: F36845

A:Status: preliminary



A:Molecule type: DNA  
A:Residues: 1-185 <BLI>  
A:Cross-references: GB:X69198; NID:g456758; PIDN:CAA49022.1; PID:g297261  
A:Experimental source: strain India-1967, ssp. major, isolate Ind3  
R:Shchelkunov, S.N.; Blinov, V.M.; Tomenin, A.V.; Marennikova, S.S.; Kolykhalov, A.A.;  
dzhaparidze, O.G.; Sandakhchiev, L.S.  
Virus Res. 27, 25-35, 1993  
A:Title: Nucleotide sequence analysis of variola virus HindIII M, L, I genome fragments.  
A:Reference number: S33069; MUID:93190624  
A:Accession: S33095  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-185 <SHC>  
A:Cross-references: EMBL:X67119; NID:g62330; PIDN:CAA47580.1; PID:g62357  
A:Experimental source: strain India-1967, isolate Ind3  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992  
C:Superfamily: vaccinia virus DNA-directed RNA polymerase 22K polypeptide  
C:Keywords: nucleotidyltransferase; transcription

Query Match 58.6%; Score 34; DB 2; Length 185;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 1 YRLIRILLRY 12  
| : | :|||  
Db 24 YAVINRNVLLRY 35

Search completed: February 7, 2000, 11:54:27  
Job time: 24337 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 00:59:54 ; Search time 63.71 Seconds  
(without alignments)  
5.625 Million cell updates/sec

Title: US-08-653-294-16  
Perfect score: 58  
Sequence: 1 YRLIRILLRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	63.8	411	1 YHS5_CAEEL	O18304 caenorhabdi
2	36	62.1	348	1 HLA_F_MACMU	P33617 macaca mula
3	35.5	61.2	278	1 YD87_YEAST	P35178 saccharomyc
4	35	60.3	298	1 RT03_ACACA	P46754 acanthamoeb
5	34	58.6	78	1 RPOH_METJA	Q58443 methanococc
6	34	58.6	185	1 DP1_MOUSE	Q60870 mus musculu
7	34	58.6	185	1 RPO6_VACCV	P07391 vaccinia vi
8	34	58.6	185	1 RPO6_VARV	P33054 variola vir
9	34	58.6	213	1 KAD_RICPR	Q92CS6 rickettsia
10	34	58.6	352	1 YK50_YEAST	P36151 saccharomyc
11	34	58.6	365	1 PE23_MOUSE	P30557 mus musculu
12	34	58.6	365	1 PE23_RAT	P34980 rattus norv
13	34	58.6	390	1 PE23_HUMAN	P43115 homo sapien
14	34	58.6	411	1 PE23_RABIT	P46069 oryctolagus
15	34	58.6	575	1 YRAM_HAEIN	P45299 haemophilus
16	34	58.6	928	1 PTP3_YEAST	P40048 saccharomyc
17	34	58.6	1780	1 POLG_MVEV	Q05769 murray vall
18	33	56.9	225	1 CBIQ_SALTY	Q05598 salmonella
19	33	56.9	238	1 LICA_MYCCA	P43044 mycoplasma
20	33	56.9	319	1 OPPE_LACLA	Q07742 lactococcus
21	33	56.9	319	1 OPPE_LACLC	P50989 lactococcus
22	33	56.9	585	1 VMD2_HUMAN	P76090 homo sapien
23	33	56.9	621	1 CRTI_CERNC	P48537 cercospora
24	33	56.9	3391	1 POLG_DEN26	P29990 d genome po
25	33	56.9	3391	1 POLG_DEN27	P29991 d genome po
26	32	55.2	211	1 Y290_METJA	Q57738 methanococc
27	32	55.2	344	1 VGLM_HSV6U	Q04630 herpes simp
28	32	55.2	344	1 VGLM_HSV62	P52449 herpes simp
29	32	55.2	550	1 SYD_MYCGE	P47282 mycoplasma
30	32	55.2	653	1 GP11_SCHPO	O14357 schizosacch
31	32	55.2	687	1 SRB4_YEAST	P32569 saccharomyc
32	32	55.2	802	1 YGN9_YEAST	P53121 saccharomyc
33	32	55.2	1075	1 RPOB_MAIZE	P16023 zea mays (m
34	32	55.2	1075	1 RPOB_ORYZA	P12091 oryza sativ

RESULT 1  
YHS5\_CAEEL  
ID YHS5\_CAEEL STANDARD; PRT; 411 AA.  
AC O18304;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HYPOTHETICAL 48.2 KD PROTEIN ZK849.5 IN CHROMOSOME I.  
GN ZK849.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA KERSHAW J.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z82095; CAB05028.1;  
DR WORMPEP; ZK849.5; CE16750.  
DR PFAM; PF01062; Worm\_family\_8; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 411 AA; 48236 MW; B5736F70 CRC32;

Query Match 63.8%; Score 37; DB 1; Length 411;  
Best Local Similarity 70.0%; Pred No. 5.8;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLIRILLRY 12  
||||| :|||  
DB 128 LLIRITIRY 137

RESULT 2  
HLAF\_MACMU  
ID HLAF\_MACMU STANDARD; PRT; 348 AA.  
AC P33617;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F  
DE ANTIGEN) (LEUKOCYTE ANTIGEN F).  
GN HLA-F OR HLA-F.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopithecinae;  
OC Macaca.

P38717 saccharomyc  
P23174 cricetus  
P34241 saccharomyc  
P45481 mus musculu  
P47760 chlamydia c  
P06796 mus musculu  
P25789 homo sapien  
P21670 rattus norv  
P19390 haemophilus  
P19391 haemophilus  
P22235 haemophilus

RN SEQUENCE FROM N.A.  
 RP MEDLINE: 93246295.  
 RA OTTING N., BONTROP R.E.;  
 RT "Characterization of the rhesus macaque (Macaca mulatta) equivalent  
 of HLA-F";  
 RL Immunogenetics 38:141-145(1993).  
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z21819; CAA79885.1; -  
 DR PIR: S29990; S29990.  
 DR HSP: P03989; IHSA.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 DR PFAM: PF00047; ig; 1.  
 DR PFAM: PF00129; MHC\_I; 1.  
 KW MHC I; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 21 BY SIMILARITY.  
 FT CHAIN 22 348 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 ALPHA CHAIN F.  
 FT DOMAIN 22 113 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 114 205 EXTRACELLULAR ALPHA-2.  
 FT DOMAIN 206 297 EXTRACELLULAR ALPHA-3.  
 FT DOMAIN 298 307 CONNECTING PEPTIDE.  
 FT TRANSMEM 308 331  
 FT DOMAIN 332 348 CYTOPLASMIC TAIL.  
 FT DISULFID 124 187 BY SIMILARITY.  
 FT DISULFID 226 282 BY SIMILARITY.  
 FT CARBOHYD 109 109 POTENTIAL.  
 SQ SEQUENCE 348 AA; 39300 MW; 3A375142 CRC32;

Query Match 62.1%; Score 36; DB 1; Length 348;  
 Best Local Similarity 54.5%; Pred. No. 7.6;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLLIRILLRY 12  
 DB 98 RVALRKLLRY 108

RESULT 3  
 YD87\_YEAST  
 ID YD87\_YEAST STANDARD; PRT; 278 AA.  
 AC P35178;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HYPOTHETICAL 33.2 KD PROTEIN IN SSSI-SL07 INTERGENIC REGION.  
 GN YD87C OR D4478.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomyces.  
 CC -----  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-S288C / FY1679;  
 RA COSTER F., JONIAUX J.-L., GOFFEAU A.;  
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 145-278 FROM N.A.  
 RX MEDLINE: 94038890.  
 RA ESNAULT Y., BLONDEL M.-O., DESHAIES R.J., SCHEKMAN R., KEPES F.;  
 RT "The yeast SSSI gene is essential for secretory protein translocation

RT and encodes a conserved protein of the endoplasmic reticulum.";  
 RL EMBL J. 12:4083-4093(1993).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE NNP-1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: Z46796; CAA86809.1; -  
 DR EMBL: Z74383; CAA98907.1; -  
 DR EMBL: X82086; CAA57616.1; -  
 DR EMBL: X74499; CAA52607.1; -  
 DR PIR: S48776; S48776.  
 KW Hypothetical protein; Nuclear protein.  
 FT DOMAIN 266 274 POLY-GLU.  
 SQ SEQUENCE 278 AA; 33202 MW; PAD7CAC9 CRC32;

Query Match 61.2%; Score 35.5; DB 1; Length 278;  
 Best Local Similarity 60.0%; Pred. No. 7.5;  
 Matches 9; Conservative 2; Mismatches 1; Indels 3; Gaps 1;

QY 1 YLLIRRL---LRY 12  
 DB 121 YLLIRRVLFSQLY 135

RESULT 4  
 RT03\_ACACA  
 ID RT03\_ACACA STANDARD; PRT; 298 AA.  
 AC P46754;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1999 (Rel. 39, Last annotation update)  
 DE MITOCHONDRIAL RIBOSOMAL PROTEIN S3.  
 GN RPS3.  
 OS Acanthamoeba castellanii (Amoeba).  
 OC Mitochondrion.  
 CC Eukaryota; Acanthamoebidae; Acanthamoeba.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 30010 / NEFF;  
 RX MEDLINE: 95147275.  
 RA BURGER G., PLANTE I., LONERGAN K.M., GRAY M.W.;  
 RT "The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba  
 castellanii: complete sequence, gene content and genome  
 organization";  
 RT J. Mol. Biol. 245:522-537(1995).  
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.  
 CC -!- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U12386; AAD11841.1; -  
 DR PROSITE: PS00548; RIBOSOMAL\_S3; FALSE\_NEG.  
 DR PFAM: PF00189; Ribosomal\_S3\_C; 1.  
 DR PFAM: PF00417; Ribosomal\_S3\_N; 1.  
 KW Ribosomal protein; Mitochondrion.  
 SQ SEQUENCE 298 AA; 36060 MW; 7BA48AD7 CRC32;

Query Match 60.3%; Score 35; DB 1; Length 298;

Best Local Similarity 50.0%; Pred. No. 10;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIRIRILLY 12  
: ||: ||: ||:  
DB 126 FMLLKRVILY 137

RESULT 5  
RPOH\_METJA  
ID RPOH\_METJA STANDARD; PRT; 78 AA.  
AC Q58443;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT H (EC 2.7.7.6).  
GN RPOH OR MJ1039.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE; 96337999.  
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,  
RA OVERBEK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GÖRGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., RAINE B.P., BORODOVSKY M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii";  
RL Science 273:1058-1073(1996).  
RN [2]  
RP STRUCTURE BY NMR.  
RX MEDLINE; 99208760.  
RA THIRU A., HODACH M., ELORANTA J.J., KOSTOUROU V., WEINZIERL R.O.,  
RA MATTHEWS S.;  
RT "RNA polymerase subunit H features a beta-ribbon motif within a novel  
fold that is present in archaea and eukaryotes";  
RL J. Mol. Biol. 287:753-760(1999).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
RNA(N).  
CC -!- SIMILARITY: TO OTHER ARCHAEABACTERIAL RPOH AND TO THE C-TERMINAL  
OF EUKARYOTIC SUBUNIT ABC27 (RPB5).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U67546; AAB99042.1; -  
CC FDB; IHMJ; 05-APR-99.  
CC TIGR; MJ1039; -  
CC PFAM; PF01191; RNA\_POL\_H:1.  
KW Transcription; Transcription; DNA-directed RNA polymerase; 3D-structure.  
SQ SEQUENCE 78 AA; 9001 MW; 9F10C0F3 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 78;  
Best Local Similarity 66.7%; Pred. No. 3.8;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLIRIRIL 9

Db 70 YRLVIKRII 78  
|||: ||: ||:

RESULT 6  
DPI\_MOUSE  
ID DPI\_MOUSE STANDARD; PRT; 185 AA.  
AC Q60870;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POLYPOSIS LOCUS PROTEIN 1 HOMOLOG (TB2 PROTEIN HOMOLOG) (GP106).  
GN DPI.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96194804.  
RA PRIESCHL E.E., PENDL G.G., HARRER N.E., BAUMRUKER T.;  
RT "The murine homolog of TB2/DPI, a gene of the familial adenomatous  
polyposis (FAP) locus";  
RL Gene 189:215-218(1996).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: TO C.ELEGANS T19C3.4.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U28168; AAB07994.1; -  
CC MGD; MGI:1270152; DPI.  
KW Transmembrane.  
FT TRANSMEM 31 51 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
SQ SEQUENCE 185 AA; 21050 MW; FC5BA4A2 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 185;  
Best Local Similarity 58.3%; Pred. No. 9.5;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRIRILLY 12  
||: ||: ||:  
DB 133 YRRIIRPIFLRH 144

RESULT 7  
RPO6\_VACCV  
ID RPO6\_VACCV STANDARD; PRT; 185 AA.  
AC P07391;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE 22 KD POLYPEPTIDE (EC 2.7.7.6).  
GN RPO22 OR J4R OR F10.  
OS Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WR;  
RX MEDLINE; 86205852.  
RA BROYLES S.S., MOSS B.;  
RT "Homology between RNA polymerases of poxviruses, prokaryotes, and  
eukaryotes: nucleotide sequence and transcriptional analysis of  
vaccinia virus genes encoding 147-kDa and 22-kDa subunits";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:3141-3145(1986).  
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN-WR;
RX MEDLINE: 85215527.
RA PLUCIENNICZAK A., SCHROEDER E., ZETTELMEISSL G., STRECK R.E.;
RT "Nucleotide sequence of a cluster of early and late genes in a
RL conserved segment of the vaccinia virus genome.";
RN Nucleic Acids Res. 13:985-998(1985).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-COPENHAGEN;
RX MEDLINE: 91021027.
RA GOBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus.";
RL virology 179:247-265(1990).
[4]
RP COMPLETE GENOME.
RC STRAIN-COPENHAGEN;
RA GOBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL virology 179:517-563(1990).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: THIS ENZYME CONSISTS OF AT LEAST EIGHT SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13209; AAB59833.1; -
DR EMBL: X01978; CAA26018.1; -
DR EMBL: M35027; AAA48084.1; -
DR PIR: I23092; RNV22.
KW Transferase; DNA-directed RNA polymerase; Transcription.
SQ SEQUENCE 185 AA; 21342 MW; 2947E2C5 CRC32;
-----
Query Match 58.6%; Score 34; DB 1; Length 185;
Best Local Similarity 50.0%; Pred. No. 9.5;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YRLIRILLRY 12
Db 24 YAVINRNVLRY 35

RESULT 8
RP06_VARV
ID RP06_VARV STANDARD; PRT; 185 AA.
AC P33054;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE 22 KD POLYPEPTIDE (EC 2.7.7.6).
GN RP022 OR J4R OR L4R.
OS Variola virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-INDIA-1967 / ISOLATE IND3;
RX MEDLINE: 93190624.
RA SHCHELKUNOV S.N., BLINOV V.M., TOTMENIN A.V., MARENNIKOVA S.S.,
RA KOLYKHAILOV A.A., FROLOV I.V., CHIZHIKOV V.E., GYTOROV V.V.,
RA GASHIKOV P.V., BELAVIN E.F., BELAVIN P.A., RESENCHUK S.M.,
RA ANDZHAPARIDZE O.G., SANDAKHCHIEV L.S.;

```

```

RT "Nucleotide sequence analysis of variola virus HindIII M, L, I genome
fragments.";
RL Virus Res. 27:25-35(1993).
[2]
RP COMPLETE GENOME.
RC STRAIN-INDIA-1967 / ISOLATE IND3;
RX MEDLINE: 93202281.
RA SHCHELKUNOV S.N., BLINOV V.M., SANDAKHCHIEV L.S.;
RT "Genes of variola and vaccinia viruses necessary to overcome the host
RT protective mechanisms.";
RL FEBS Lett. 319:80-83(1993).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: THIS ENZYME CONSISTS OF AT LEAST EIGHT SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67119; CAA47580.1; -
DR EMBL: S5844; AAB4677.1; -
DR EMBL: X69198; CAA49022.1; -
DR PIR: F36845; F36845.
DR PIR: S33095; S33095.
KW Transferase; DNA-directed RNA polymerase; Transcription.
SQ SEQUENCE 185 AA; 21354 MW; 547A4AC2 CRC32;
-----
Query Match 58.6%; Score 34; DB 1; Length 185;
Best Local Similarity 50.0%; Pred. No. 9.5;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YRLIRILLRY 12
Db 24 YAVINRNVLRY 35

RESULT 9
KAD_RICPR
ID KAD_RICPR STANDARD; PRT; 213 AA.
AC Q9ZCS6;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN ADK OR Rp638.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RX MEDLINE: 99039499.
RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA SICKERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA ERIKSSON A.-S., WINKLER H.H., KURLAND C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -!- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AJ235272; CAA15078.1; -  
DR PROSITE; PS00113; ADENYLATE KINASE; 1.  
KW Transferase; Kinase; ATP-binding.  
FT NP\_BIND 7 15 ATP (BY SIMILARITY).  
SQ SEQUENCE 213 AA; 24503 MW; 3CFD3026 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 213;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLIRRLRYL 12  
Db 116 LLIRRLRYL 125

RESULT 10  
YK50 YEAST  
ID YK50 YEAST STANDARD; PRT; 352 AA.  
AC P36151;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 39.4 KD PROTEIN IN MET1-SIS2 INTERGENIC REGION.  
GN YK070W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA POHL T.M., POHL F.M.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z28295; CAA82149.1; -  
DR PIR; S38147; S38147.  
KW Hypothetical protein.  
SQ SEQUENCE 352 AA; 39406 MW; F5B1BB2A CRC32;

Query Match 58.6%; Score 34; DB 1; Length 352;  
Best Local Similarity 60.0%; Pred. No. 19;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YLLIRRL 10  
Db 223 FLLVRLRYL 232

RESULT 11  
PE23 MOUSE  
ID PE23 MOUSE STANDARD; PRT; 365 AA.  
AC P30557;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE PROTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE  
DE RECEPTOR, EP3 SUBTYPE).

GN PTGER3 OR PTGEREP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A. (ALPHA FORM).  
RC STRAIN-BDF1;  
RA MEDLINE; 92202182.  
RX SUGIMOTO Y., NAMBA T., HONDA A., HAYASHI Y., NEGISHI M.,  
RA ICHIKAWA A., NARUMIYA S.;  
RT "Cloning and expression of a cDNA for mouse prostaglandin E receptor  
EP3 subtype.";  
RL J. Biol. Chem. 267:6463-6466(1992).  
RN [2]  
RP SEQUENCE FROM N.A. (BETA FORM).  
RX MEDLINE; 93155085.  
RA SUGIMOTO Y., NEGISHI M., HAYASHI Y., NAMBA T., HONDA A., WATABE A.,  
RA HIRATA M., NARUMIYA S., ICHIKAWA A.;  
RT "Two isoforms of the EP3 receptor with different carboxyl-terminal  
domains. Identical ligand binding properties and different coupling  
properties with Gi proteins.";  
RL J. Biol. Chem. 268:2712-2718(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (GAMMA FORM).  
RX MEDLINE; 94039052.  
RA IRIE A., SUGIMOTO Y., NAMBA T., HARAZONO A., HONDA A., WATABE A.,  
RA NEGISHI M., NARUMIYA S., ICHIKAWA A.;  
RT "Third isoform of the prostaglandin-E-receptor EP3 subtype with  
different C-terminal tail coupling to both stimulation and inhibition  
of adenylate cyclase.";  
RL Eur. J. Biochem. 217:313-318(1993).  
RN [4]  
RP CHARACTERIZATION (ALPHA AND BETA FORMS).  
RX MEDLINE; 96147152.  
RA HASEGAWA H., NEGISHI M., ICHIKAWA A.;  
RT "Two isoforms of the prostaglandin E receptor EP3 subtype different  
in agonist-independent constitutive activity.";  
RL J. Biol. Chem. 271:1857-1860(1996).  
CC -!- FUNCTION: RECEPTOR FOR PROTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR  
CC MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION.  
CC MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL  
CC NEURONS, INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY  
CC TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF  
CC THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE  
CC CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF  
CC INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH  
CC DIFFERENT SECOND MESSENGER SYSTEMS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST THREE ISOFORMS; ALPHA (SHOWN HERE),  
CC BETA AND GAMMA; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE  
CC IDENTICAL LIGAND BINDING PROPERTIES BUT DIFFERENT COUPLING  
CC PROPERTIES WITH G PROTEINS: ALPHA AND BETA COUPLE TO G-I PROTEINS,  
CC WHEREAS GAMMA COUPLES TO MULTIPLE G-PROTEINS, G-I AND G-S. ALPHA  
CC SHOWS HIGH AGONIST-INDEPENDENT CONSTITUTIVE INHIBITION OF  
CC ADENYLATE CYCLASE, WHILE BETA HAS NO AGONIST INDEPENDENT  
CC INHIBITION.  
CC -!- TISSUE SPECIFICITY: KIDNEY, UTERUS, AND MASTOCYTOCYTES, AND  
CC IN A LESSER AMOUNT IN BRAIN, THYMUS, LUNG, HEART, STOMACH, AND  
CC SPLEEN.  
CC -!- PTM: LIGAND BINDING IS AFFECTED BY CAMP-DEPENDENT PHOSPHORYLATION  
CC IN BRAIN MEMBRANES.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D10204; BAA01051.1; -  
DR EMBL; D13321; BAA02578.1; -





FT DOMAIN 131 151 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 152 173 4 (POTENTIAL).  
 FT DOMAIN 174 203 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 204 229 5 (POTENTIAL).  
 FT DOMAIN 230 259 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 260 283 6 (POTENTIAL).  
 FT DOMAIN 284 303 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 304 325 7 (POTENTIAL).  
 FT DOMAIN 326 355 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 16 16 POTENTIAL.  
 FT CARBOHYD 193 193 POTENTIAL.  
 FT DISULFID 107 184 BY SIMILARITY.  
 FT VARSPLIC 336 365 IDHTNYASSSTSLPCGSSVLMWSLDOLR -> MNMLKLR  
 FT VARSPLIC 336 365 SFIAPASLSMRISPPREG (IN ISOFORM BETA).  
 FT IDHTNYASSSTSLPCGSSVLMWSLDOLR -> VANAYSS  
 FT CSDDQKGQALSLSENVHPGP (IN ISOFORM GAMMA).  
 FT VARIANT 152 152 P -> RA.  
 FT CONFLICT 51 51 V -> S (IN REF. 3 AND 4).  
 FT CONFLICT 354 354 S -> F (IN REF. 3).  
 FT SEQUENCE 365 AA; 39942 MW; F302B36A CRC32;  
 SQ SEQUENCE 365 AA; 39942 MW; F302B36A CRC32;  
 Query Match 58.6%; Score 34; DB 1; Length 365;  
 Best Local Similarity 77.8%; Pred. No. 19;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 LLIRILLR 11  
 DB 323 LLIRKILLR 331  
 RESULT 13  
 PE23\_HUMAN STANDARD; PRT; 390 AA.  
 AC P43115; Q12943; Q12944; Q12945; Q16546;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE  
 DE RECEPTOR, EP3 SUBTYPE).  
 GN PTGER3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94139918.  
 RA ADAM M., BOIE Y., RUSHMORE T.H., MULLER G., BASTIEN L., MCKEE K.T.,  
 RA METTERS K.M., ABRAMOVITZ M.;  
 RT "Cloning and expression of three isoforms of the human EP3 prostanoid  
 RT receptor.";  
 RL FEBS Lett. 338:170-174(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-UTERUS;  
 RX MEDLINE; 95188908.  
 RA SCHMID A., THERRAUCH K.H., SCHLEUNING W.D., DINTER H.;  
 RT "Splice variants of the human EP3 receptor for prostaglandin E2.";  
 RL Eur. J. Biochem. 228:23-30(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE; 94161771.  
 RA YANG J., XIA M., GOETZL E.J., AN S.;  
 RT "Cloning and expression of the EP3-subtype of human receptors for  
 RT prostaglandin E2.";  
 RL Biochem. Biophys. Res. Commun. 198:999-1006(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94183149.  
 RA KUNAPULI S.P., FEN MAO G., BASTEPE M., LIU-CHEN L.-Y., LI S.,  
 RA CHEUNG P.P., DERIEL J.K., ASHBY B.;

RT "Cloning and expression of a prostaglandin E receptor EP3 subtype  
 from human erythroleukemia cells.";  
 RL Biochem. J. 298:263-267(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-UTERUS;  
 RC MEDLINE; 95072021.  
 RA AN S., YANG J., SO S.W., ZENG L., GOETZL E.J.;  
 RT "Isoforms of the EP3 subtype of human prostaglandin E2 receptor  
 RT transduce both intracellular calcium and cAMP signals.";  
 RL Biochemistry 33:14496-14502(1994).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE; 96074809.  
 RA KOTANI M., TANAKA I., OGAWA Y., USUI T., MORI K., ICHIKAWA A.,  
 RA NARUMIYA S., YOSHIMI T., NAKAO K.;  
 RT "Molecular cloning and expression of multiple isoforms of human  
 RT prostaglandin E receptor EP3 subtype generated by alternative  
 RT messenger RNA splicing: multiple second messenger systems and  
 RT tissue-specific distributions.";  
 RL Mol. Pharmacol. 48:869-879(1995).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SMALL INTESTINE;  
 RX MEDLINE; 94356288.  
 RA REGAN J.W., BAILEY T.J., DONELLO J.E., PIERCE K.L., PEPPERL D.J.,  
 RA ZHANG D., KEDZIE K.M., FAIRBAIRN C.E., BOGARDUS A.M., WOODWARD D.F.,  
 RA GIL D.W.;  
 RT "Molecular cloning and expression of human EP3 receptors: evidence of  
 RT three variants with differing carboxyl termini.";  
 RL Br. J. Pharmacol. 112:377-385(1994).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97230456.  
 RA KOTANI M., TANAKA I., OGAWA Y., USUI T., TAMURA N., MORI K.,  
 RA NARUMIYA S., YOSHIMI T., NAKAO K.;  
 RT "Structural organization of the human prostaglandin EP3 receptor  
 RT subtype gene (PTGER3).";  
 RL Genomics 40:425-434(1997).  
 CC -!- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR  
 CC MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION,  
 CC MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL  
 CC NEURONS, INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY  
 CC TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF  
 CC THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE  
 CC CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF  
 CC INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH  
 CC DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST SIX FORMS THAT DIFFERS AT THE C-  
 CC TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS RECEPTOR  
 CC GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT DIFFERENT  
 CC COUPLING PROPERTIES WITH G PROTEINS. THE FORMS SHOWN HERE IS KNOWN  
 CC AS EP3-1, EP3A, OR EP3AL.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT LEAST IN SMALL INTESTINE, HEART,  
 CC AND PANCREAS.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -!- CAUTION: IN REF.7 EP3B IS KNOWN AS EP3E, EP3C AS EP3D AND EP3D AS  
 CC EP3F.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S69200; AAC29854.1; -;  
 DR EMBL; L27490; AAC13374.1; -;  
 DR EMBL; L27488; AAC13372.1; -;  
 DR EMBL; L27489; AAC13373.1; -;

DR	EMBL	X83857	CAA58737.1	-	FT	TRANSEM	228	253	5 (POTENTIAL).
DR	EMBL	X83858	CAA58738.1	-	FT	DOMAIN	254	283	CYTOPLASMIC (POTENTIAL).
DR	EMBL	X83859	CAA58739.1	-	FT	TRANSEM	284	307	6 (POTENTIAL).
DR	EMBL	X83860	CAA58740.1	-	FT	DOMAIN	308	327	EXTRACELLULAR (POTENTIAL).
DR	EMBL	X83861	CAA58741.1	-	FT	TRANSEM	328	349	7 (POTENTIAL).
DR	EMBL	X83862	CAA58742.1	-	FT	DOMAIN	350	390	CYTOPLASMIC (POTENTIAL).
DR	EMBL	X83863	CAA58743.1	-	FT	CARBOHYD	18	18	POTENTIAL.
DR	EMBL	L26976	AAA60076.1	-	FT	CARBOHYD	36	36	POTENTIAL.
DR	EMBL	S69326	AAA30208.1	-	FT	VARSPLIC	360	390	IRYHTNNYASSSTSLPQCSCSTLWMSDHLER -> VANAVS
DR	EMBL	L32660	AAA68191.1	-	FT	EMBL			SCSNDGKGQIPISLSNIIQTEA (IN ISOFORM
DR	EMBL	L32661	AAA68192.1	-	FT	EMBL			EP3C/EP3-II).
DR	EMBL	L32662	AAA68193.1	-	FT	VARSPLIC	360	390	IRYHTNNYASSSTSLPQCSCSTLWMSDHLER -> BEFWGN
DR	EMBL	D38297	BAAO7416.1	-	FT	EMBL			(IN ISOFORM EP3B/EP3-III).
DR	EMBL	D38298	BAAO7417.1	-	FT	VARSPLIC	360	390	IRYHTNNYASSSTSLPQCSCSTLWMSDHLER -> MKRRL
DR	EMBL	D38299	BAAO7418.1	-	FT	EMBL			REQEEFWGN (IN ISOFORM EP3D).
DR	EMBL	D38300	BAAO7419.1	-	FT	EMBL			
DR	EMBL	D38301	BAAO7420.1	-	FT	EMBL			
DR	EMBL	U13214	AAA21130.1	-	FT	EMBL			
DR	EMBL	U13215	AAA21131.1	-	FT	EMBL			
DR	EMBL	U13216	AAA21132.1	-	FT	EMBL			
DR	EMBL	U13217	AAA21133.1	-	FT	EMBL			
DR	EMBL	U13218	AAA21134.1	-	FT	EMBL			
DR	EMBL	D86096	BAAL19951.1	-	FT	EMBL			
DR	EMBL	D86087	BAAL19951.1	JOINED.	FT	EMBL			
DR	EMBL	D86088	BAAL19951.1	JOINED.	FT	EMBL			
DR	EMBL	GCR_0837	-	-	DR	GCRDB	GCR_0837	-	-
DR	EMBL	GCR_0886	-	-	DR	GCRDB	GCR_0886	-	-
DR	EMBL	GCR_0887	-	-	DR	GCRDB	GCR_0887	-	-
DR	EMBL	GCR_0888	-	-	DR	GCRDB	GCR_0888	-	-
DR	EMBL	GCR_1279	-	-	DR	GCRDB	GCR_1279	-	-
DR	EMBL	GCR_1288	-	-	DR	GCRDB	GCR_1288	-	-
DR	EMBL	GCR_1809	-	-	DR	GCRDB	GCR_1809	-	-
DR	EMBL	GCR_1810	-	-	DR	GCRDB	GCR_1810	-	-
DR	EMBL	GCR_1811	-	-	DR	GCRDB	GCR_1811	-	-
DR	EMBL	GCR_1812	-	-	DR	GCRDB	GCR_1812	-	-
DR	EMBL	GCR_1813	-	-	DR	GCRDB	GCR_1813	-	-
DR	EMBL	GCR_1866	-	-	DR	GCRDB	GCR_1866	-	-
DR	EMBL	GCR_1867	-	-	DR	GCRDB	GCR_1867	-	-
DR	EMBL	GCR_1868	-	-	DR	GCRDB	GCR_1868	-	-
DR	EMBL	GCR_2037	-	-	DR	GCRDB	GCR_2037	-	-
DR	EMBL	GCR_2038	-	-	DR	GCRDB	GCR_2038	-	-
DR	EMBL	GCR_2039	-	-	DR	GCRDB	GCR_2039	-	-
DR	EMBL	GCR_2041	-	-	DR	GCRDB	GCR_2041	-	-
DR	EMBL	GCR_2099	-	-	DR	GCRDB	GCR_2099	-	-
DR	EMBL	GCR_2100	-	-	DR	GCRDB	GCR_2100	-	-
DR	EMBL	GCR_2101	-	-	DR	GCRDB	GCR_2101	-	-
DR	EMBL	GCR_2102	-	-	DR	GCRDB	GCR_2102	-	-
DR	EMBL	GCR_2103	-	-	DR	GCRDB	GCR_2103	-	-
DR	EMBL	GCR_2104	-	-	DR	GCRDB	GCR_2104	-	-
DR	EMBL	GCR_2105	-	-	DR	GCRDB	GCR_2105	-	-
DR	EMBL	GCR_2660	-	-	DR	GCRDB	GCR_2660	-	-
DR	EMBL	GCR_2661	-	-	DR	GCRDB	GCR_2661	-	-
DR	EM								

DR	EMBL	X83857	CAA58737.1	-	FT	TRANSEM	228	253	5 (POTENTIAL).
DR	EMBL	X83858	CAA58738.1	-	FT	DOMAIN	254	283	CYTOPLASMIC (POTENTIAL).
DR	EMBL	X83859	CAA58739.1	-	FT	TRANSEM	284	307	6 (POTENTIAL).
DR	EMBL	X83860	CAA58740.1	-	FT	DOMAIN	308	327	EXTRACELLULAR (POTENTIAL).
DR	EMBL	X83861	CAA58741.1	-	FT	TRANSEM	328	349	7 (POTENTIAL).
DR	EMBL	X83862	CAA58742.1	-	FT	DOMAIN	350	390	CYTOPLASMIC (POTENTIAL).
DR	EMBL	X83863	CAA58743.1	-	FT	CARBOHYD	18	18	POTENTIAL.
DR	EMBL	L26976	AAA60076.1	-	FT	CARBOHYD	36	36	POTENTIAL.
DR	EMBL	S69326	AAA30208.1	-	FT	VARSPLIC	360	390	IRYHTNNYASSSTSLPQCSCSTLMWSDHLER -> VANAVS
DR	EMBL	L32660	AAA68191.1	-	FT	EMBL			SCSNDGKGQIPISLSNIIQTEA (IN ISOFORM
DR	EMBL	L32661	AAA68192.1	-	FT	EMBL			EP3C/EP3-II).
DR	EMBL	L32662	AAA68193.1	-	FT	VARSPLIC	360	390	IRYHTNNYASSSTSLPQCSCSTLMWSDHLER -> BEFWGN
DR	EMBL	D38297	BAAO7416.1	-	FT	EMBL			(IN ISOFORM EP3B/EP3-III).
DR	EMBL	D38298	BAAO7417.1	-	FT	VARSPLIC	360	390	IRYHTNNYASSSTSLPQCSCSTLMWSDHLER -> MKRRL
DR	EMBL	D38299	BAAO7418.1	-	FT	EMBL			REQEEFWGN (IN ISOFORM EP3D).
DR	EMBL	D38300	BAAO7419.1	-	FT	EMBL			
DR	EMBL	D38301	BAAO7420.1	-	FT	EMBL			
DR	EMBL	U13214	AAA21130.1	-	FT	EMBL			
DR	EMBL	U13215	AAA21131.1	-	FT	EMBL			
DR	EMBL	U13216	AAA21132.1	-	FT	EMBL			
DR	EMBL	U13217	AAA21133.1	-	FT	EMBL			
DR	EMBL	U13218	AAA21134.1	-	FT	EMBL			
DR	EMBL	D86096	BAAL19951.1	-	FT	EMBL			
DR	EMBL	D86087	BAAL19951.1	JOINED.	FT	EMBL			
DR	EMBL	D86088	BAAL19951.1	JOINED.	FT	EMBL			
DR	EMBL	GCR_0837	-	-	DR	GCRDB	GCR_0837	-	-
DR	EMBL	GCR_0886	-	-	DR	GCRDB	GCR_0886	-	-
DR	EMBL	GCR_0887	-	-	DR	GCRDB	GCR_0887	-	-
DR	EMBL	GCR_0888	-	-	DR	GCRDB	GCR_0888	-	-
DR	EMBL	GCR_1279	-	-	DR	GCRDB	GCR_1279	-	-
DR	EMBL	GCR_1288	-	-	DR	GCRDB	GCR_1288	-	-
DR	EMBL	GCR_1809	-	-	DR	GCRDB	GCR_1809	-	-
DR	EMBL	GCR_1810	-	-	DR	GCRDB	GCR_1810	-	-
DR	EMBL	GCR_1811	-	-	DR	GCRDB	GCR_1811	-	-
DR	EMBL	GCR_1812	-	-	DR	GCRDB	GCR_1812	-	-
DR	EMBL	GCR_1813	-	-	DR	GCRDB	GCR_1813	-	-
DR	EMBL	GCR_1866	-	-	DR	GCRDB	GCR_1866	-	-
DR	EMBL	GCR_1867	-	-	DR	GCRDB	GCR_1867	-	-
DR	EMBL	GCR_1868	-	-	DR	GCRDB	GCR_1868	-	-
DR	EMBL	GCR_2037	-	-	DR	GCRDB	GCR_2037	-	-
DR	EMBL	GCR_2038	-	-	DR	GCRDB	GCR_2038	-	-
DR	EMBL	GCR_2039	-	-	DR	GCRDB	GCR_2039	-	-
DR	EMBL	GCR_2041	-	-	DR	GCRDB	GCR_2041	-	-
DR	EMBL	GCR_2099	-	-	DR	GCRDB	GCR_2099	-	-
DR	EMBL	GCR_2100	-	-	DR	GCRDB	GCR_2100	-	-
DR	EMBL	GCR_2101	-	-	DR	GCRDB	GCR_2101	-	-
DR	EMBL	GCR_2102	-	-	DR	GCRDB	GCR_2102	-	-
DR	EMBL	GCR_2103	-	-	DR	GCRDB	GCR_2103	-	-
DR	EMBL	GCR_2104	-	-	DR	GCRDB	GCR_2104	-	-
DR	EMBL	GCR_2105	-	-	DR	GCRDB	GCR_2105	-	-
DR	EMBL	GCR_2660	-	-	DR	GCRDB	GCR_2660	-	-
DR	EMBL	GCR_2661	-	-	DR	GCRDB	GCR_2661	-	-
DR	EMBL								

```

DR EMBL; U04275; AAA17414.1; -.
DR EMBL; U04276; AAA17415.1; -.
DR GCRDB; GCR_0954; -.
DR GCRDB; GCR_0955; -.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
DR PFAM; PF00001; 7tm_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Alternative splicing.
FT DOMAIN 1 49 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 50 74 1 (POTENTIAL).
FT DOMAIN 75 87 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 88 108 2 (POTENTIAL).
FT DOMAIN 109 127 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 128 149 3 (POTENTIAL).
FT DOMAIN 150 170 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 172 193 4 (POTENTIAL).
FT DOMAIN 194 223 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 224 249 5 (POTENTIAL).
FT DOMAIN 250 279 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 280 303 6 (POTENTIAL).
FT DOMAIN 304 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 345 7 (POTENTIAL).
FT DOMAIN 346 411 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 18 18 POTENTIAL.
FT CARBOHYD 32 32 POTENTIAL.
FT CARBOHYD 213 213 BY SIMILARITY.
FT DISULFID 126 204 T -> A (IN CLONE 72A).
FT VARIANT 243 243 VIHENN -> EEFWEK (IN CLONE 72A).
FT VARSPLIC 356 361 MISSING (IN CLONE 72A).
FT VARSPLIC 362 411 VIHENNEKDIQRENVSHS -> IRYHTNNYASSTSL
FT VARSPLIC 356 377 THQCSST (IN CLONE 74A).
FT VARSPLIC 378 411 MISSING (IN CLONE 74A).
FT VARSPLIC 356 389 VIHENNEKDIQRENVSHSQHEARDESEK -> HSP
FT AIGLQISTHISKNKYFGLMKYTHSLAYL (IN CLONE
FT 80A).
FT VARSPLIC 390 411 MISSING (IN CLONE 80A).
FT VARSPLIC 411 AA; 45600 MW; 299918C2 CRC32;
SQ SEQUENCE

```

Query Match 58.68; Score 34; DB 1; Length 411;  
 Best Local Similarity 77.8%; Pred. No. 22;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 LLIRRIILLR 11  
 ||:|||||  
 Db 343 LLRLKILLR 351

```

RESULT 15
YRAM_HAEIN STANDARD; PRT; 575 AA.
AC P45299;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL PROTEIN H11655.
GN H11655.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,

```

```

RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SIMILARITY: STRONG, TO E.COLI YRAM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32838; AAC23299.1; -.
DR TIGR; H11655; -.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 575 AA; 63436 MW; 90038AEB CRC32;

Query Match 58.68; Score 34; DB 1; Length 575;
Best Local Similarity 54.5%; Pred. No. 31;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 YRLIRILLR 11
|:|:|:|:|
Db 65 YKLLAARVLIR 75

Search completed: February 8, 2000, 00:59:55
Job time: 3784 sec

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 13:17:42 ; Search time 209.03 Seconds  
(without alignments)  
3.980 Million cell updates/sec

Title: US-08-653-294-16  
Perfect score: 58  
Sequence: 1 YRLIRILLRY 12

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

SPTREMBL12:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.rodent:\*
- 12: sp.virus:\*
- 13: sp.vertebrate:\*
- 14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	65.5	260	12 Q9YW71	Q9YW71 melanoplus
2	37	63.8	408	2 O66951	O66951 aquifex aeo
3	36	62.1	559	2 O84647	O84647 chlamydia t
4	36	62.1	1054	4 O94887	O94887 homo sapien
5	36	62.1	1055	2 O05052	O05052 haemophilus
6	35	60.3	799	5 Q27591	Q27591 drosophila
7	35	60.3	814	10 O64487	O64487 arabidopsis
8	35	60.3	1712	5 O96160	O96160 plasmodium
9	35	60.3	1847	5 P91495	P91495 caenorhabdi
10	34	58.6	179	12 Q85331	Q85331 vaccinia vi
11	34	58.6	185	12 Q85383	Q85383 variola vir
12	34	58.6	313	2 P97213	P97213 clostridium
13	34	58.6	325	2 Q9XB05	Q9XB05 myxococcus
14	34	58.6	393	4 O00326	O00326 homo sapien
15	34	58.6	402	4 O00325	O00325 homo sapien
16	34	58.6	454	4 O45191	O45191 homo sapien
17	34	58.6	683	5 Q9Y136	Q9Y136 drosophila
18	34	58.6	735	5 O76912	O76912 drosophila
19	34	58.6	777	4 O43162	O43162 homo sapien
20	34	58.6	2154	10 Q9ZVC8	Q9ZVC8 arabidopsis

21	33	56.9	111	2 P70780	P70780 anabaena sp
22	33	56.9	133	2 Q92903	Q92903 chlamydia p
23	33	56.9	233	1 O26858	O26858 methanobact
24	33	56.9	255	2 Q45433	Q45433 bacillus st
25	33	56.9	302	2 Q9X9Y3	Q9X9Y3 streptomyce
26	33	56.9	364	2 P97084	P97084 salmonella
27	33	56.9	425	1 O26797	O26797 methanobact
28	33	56.9	532	2 P72007	P72007 mycobacteri
29	33	56.9	728	4 Q9Y219	Q9Y219 homo sapien
30	33	56.9	795	5 Q22063	Q22063 caenorhabdi
31	33	56.9	802	3 Q92385	Q92385 tolypocladi
32	33	56.9	815	3 O59744	O59744 schizosacch
33	33	56.9	839	12 O11875	O11875 dengue viru
34	33	56.9	3391	12 O09234	O09234 unidentified
35	32	55.2	132	2 Q9X378	Q9X378 bacillus an
36	32	55.2	193	8 Q04936	Q04936 oryza sativ
37	32	55.2	217	10 Q9ZRD8	Q9ZRD8 glycine max
38	32	55.2	237	2 P72295	P72295 rhizobium m
39	32	55.2	258	3 Q06790	Q06790 saccharomyc
40	32	55.2	259	11 Q60503	Q60503 cricetus
41	32	55.2	260	5 O96917	O96917 anopheles g
42	32	55.2	263	2 Q56779	Q56779 xanthomonas
43	32	55.2	295	2 Q48733	Q48733 lactobacill
44	32	55.2	321	5 O45795	O45795 caenorhabdi
45	32	55.2	361	5 Q19771	Q19771 caenorhabdi

#### ALIGNMENTS

RESULT 1

ID Q9YW71 PRELIMINARY; PRT; 260 AA.  
AC Q9YW71;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE ORF MSV021 MTG MOTIF GENE FAMILY PROTEIN.  
GN MSV021.  
OS Melanoplus sanguinipes entomopoxvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
OC Entomopoxvirus B.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TUCSON;  
RX MEDLINE; 99102612.  
RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;  
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";  
RL J. Virol. 73:533-552(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TUCSON;  
RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF063866; AAC97850.1; -  
SQ SEQUENCE 260 AA; 31180 MW; AB880E14 CRC32;

Query Match 65.5% Score 38; DB 12; Length 260;

Best Local Similarity 66.7%; Pred. No. 14;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRILLRY 12  
| | | | | | | |  
DB 234 YLLIEKILLNY 245

RESULT 2

ID O66951 PRELIMINARY; PRT; 408 AA.  
AC O66951;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
 DE FIMBRIAL ASSEMBLY PROTEIN PILC.  
 GN PILC.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RX MEDLINE; 98196666.  
 RA GRAHAM D.E., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.,  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus."  
 RL Nature 392:353-358(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.,  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE000705; AAC06902.1;  
 DR PFAM; PF00482; GSPILF.1;  
 SQ SEQUENCE 408 AA; 45763 MW; 2E2A7221 CRC32;

Query Match 63.8%; Score 37; DB 2; Length 408;  
 Best Local Similarity 72.7%; Pred. No. 33;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 YRLIRILLR 11  
 : | | | | |  
 Db 250 FRALIRILLR 260

RESULT 3  
 ID 084647 PRELIMINARY; PRT; 559 AA.  
 AC 084647;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DE 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE EFFLUX PROTEIN.  
 GN YGED.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/UW-3/CX;  
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,  
 RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,  
 RA DAVIS R.W.;  
 RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:  
 Chlamydia trachomatis."  
 RL Science 0:0-0(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/UW-3/CX;  
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,  
 RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,  
 RA DAVIS R.W.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE001334; AAC68245.1;  
 SQ SEQUENCE 559 AA; 61931 MW; 618B74E1 CRC32;

Query Match 62.1%; Score 36; DB 2; Length 559;  
 Best Local Similarity 63.6%; Pred. No. 68;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 YRLIRILLR 11  
 : | | | | |

Db 409 YRLIRILLR 419

RESULT 4  
 ID 094887 PRELIMINARY; PRT; 1054 AA.  
 AC 094887;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE KIAA0793 PROTEIN.  
 GN KIAA0793.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 99087487.  
 RA NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., MIYAJIMA N., TANAKA A.,  
 RA KOTANI H., NOMURA N., OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XI.  
 The complete sequences of 100 new cDNA clones from brain which code  
 for large proteins in vitro."  
 RL DNA Res. 5:277-286(1998).  
 RN [2]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RX MEDLINE; 99063792.  
 RA SULSTON J.E., WATERSTON R.;  
 RT "Toward a complete human genome sequence."  
 RL Genome Res. 8:1097-1108(1998).  
 RN [3]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA SUN H., STONEKING T., LANGSTON Y., LAPLANT Y.;  
 RT "The sequence of Homo sapiens BAC clone RG442F18."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA WATERSTON R.H.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA WATERSTON R.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA WATERSTON R.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB018336; BAA34513.1;  
 DR EMBL; AC005104; AAD12224.1;  
 DR HSP; P08567; IPLS.  
 SQ SEQUENCE 1054 AA; 119888 MW; 3BA89171 CRC32;

Query Match 62.1%; Score 36; DB 4; Length 1054;  
 Best Local Similarity 58.3%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YRLIRILLR 12  
 : | | | | |  
 Db 688 YRLIRILLR 699

RESULT 5  
 ID 005052 PRELIMINARY; PRT; 1055 AA.  
 AC 005052;  
 DT 01-JUL-1997 (Tremblrel. 04, Created)  
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE TYPE I RESTRICTION ENZYME ECOR124/3 R PROTEIN (HSDR).  
 GN H1285.  
 OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 95350630.  
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
 RA KERLAVAGE A.R., BULT C.J., TOMB J.F., DOUGHERTY B.A., MERRICK J.M.,  
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
 RA SCOTT J.D., SHIRLEY R., LIU L.I., GLODEK A., KELLEY J.M.,  
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOHAGEN N.S.M.,  
 RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
 RA VENTER J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 96398784.  
 RA L TATSOV R., MUSHEGIAN A.R., BORK P., BROWN N.P., HAYES W.S.,  
 RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;  
 RT "Metabolism and evolution of Haemophilus influenzae deduced from a  
 whole-genome comparison with Escherichia coli.";  
 RL Curr. Biol. 6:279-291(1996).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
 RA KERLAVAGE A.R., BULT C.J., TOMB J.F., DOUGHERTY B.A., MERRICK J.M.,  
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
 RA SCOTT J.D., SHIRLEY R., LIU L.I., GLODEK A., KELLEY J.M.,  
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOHAGEN N.S.M.,  
 RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
 RA VENTER J.C.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RA WHITE O., CLAYTON R.A., KERLAVAGE A.R., FLEISCHMANN R.D.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U32808; AAC22934.1; -;  
 DR TIGR; H1285; -;  
 SQ SEQUENCE 1055 AA; 120709 MW; 9F1C0713 CRC32;  
 Query Match 62.1%; Score 36; DB 2; Length 1055;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 RLLIRILLRY 12  
 Db |||:||||:  
 Db 1017 RLLVRRALQY 1027  
 RESULT 6  
 ID Q27591 PRELIMINARY; PRT; 799 AA.  
 AC Q27591;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE INTEGRIN BETA SUBUNIT (FRAGMENT).  
 GN <BETA>INT.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-MIDGUT ENDODERM;  
 RX MEDLINE; 94357079.  
 RA YEE G.H., HYNES R.O.;

"A novel, tissue-specific integrin subunit, beta nu, expressed in the  
 midgut of Drosophila melanogaster.";  
 RL Development 118:845-858(1993).  
 CC -|- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE  
 CC GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX ADHESION.  
 CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -|- SIMILARITY: WITH OTHER BETA CHAINS FROM THE INTEGRIN FAMILY OF  
 CELL-SURFACE RECEPTOR.  
 DR EMBL; L13305; AAC37169.1; -;  
 DR HSSP; P00750; ITPG.  
 DR FLYBASE; Fggn0010395; beta-Int-nu.  
 DR PROSITE; PS00243; INTEGRIN\_BETA; 1.  
 DR PFAM; PF00362; Integrin\_P; 1.  
 KW Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat.  
 FT NON\_TER 1 799  
 FT NON\_TER 799 799  
 SQ SEQUENCE 799 AA; 90741 MW; 0F909972 CRC32;  
 Query Match 60.3%; Score 35; DB 5; Length 799;  
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 YRLIRILLRY 12  
 Db |||:||||:  
 Db 61 YRCLSRQLLNY 72  
 RESULT 7  
 ID Q64487 PRELIMINARY; PRT; 814 AA.  
 AC Q64487;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE F20D22.3 PROTEIN.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-CV. COLUMBIA;  
 RC VYOTSKAIA V.S., OSBORNE B.I., SCHWARTZ J.R., TORIUMI M., KWAN A.,  
 RA YU G., OUI, O., LIU S., LI J., HOANG L., ARAUJO R., AU M., BRENDEN V.,  
 RA BUEHLER E., CONWAY A.B., CONWAY A.R., DEWAR K., FENG J., KIM C.,  
 RA KURTZ D., LI Y., PALM C.J., SHINN P., SUN H., DAVIS R.W., ECKER J.R.,  
 RA FEDERSPIEL N.A., THEOLOGIS A.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-CV. COLUMBIA;  
 RC THEOLOGIS A.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-CV. COLUMBIA;  
 RC THEOLOGIS A.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-CV. COLUMBIA;  
 RC THEOLOGIS;  
 RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC002411; AAC16741.1; -;  
 SQ SEQUENCE 814 AA; 92446 MW; AE8AF982 CRC32;  
 Query Match 60.3%; Score 35; DB 10; Length 814;  
 Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLIRIRILLRY 12  
 DB 263 LVVRRLLNRY 272

RESULT 8  
 O96160 ID Q96160 PRELIMINARY; PRT; 1712 AA.  
 AC Q96160;  
 DT 01-MAY-1999 (TREMREL. 10, Created)  
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMREL. 10, Last annotation update)  
 DE HYPOTHETICAL 204.0 KD PROTEIN.  
 GN PFB0315W.  
 GN Plasmodium falciparum.  
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 99021743.  
 RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L., KOONIN E.V., SHALLOO S., MASON T., YU K., FUJII C., PEDERSON J., SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERIEA M., RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O., RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum";  
 RL Science 282:1126-1132(1998).  
 DR EMBL: AE001386; AAC71852.1; -.  
 SQ SEQUENCE 1712 AA; 204013 MW; E8B0BD71 CRC32;

Query Match 60.3%; Score 35; DB 5; Length 1712;  
 Best Local Similarity 50.0%; Pred. No. 3et02;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRILLRY 12  
 DB 1316 YKIVRLIFGY 1327

RESULT 9  
 P91495 ID P91495 PRELIMINARY; PRT; 1847 AA.  
 AC P91495;  
 DT 01-MAY-1997 (TREMREL. 03, Created)  
 DT 01-MAY-1997 (TREMREL. 03, Last sequence update)  
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
 DE SIMILARITY TO RAT INTEGRAL MEMBRANE GLYCOPROTEIN GPI20 PRECURSOR.  
 GN T23H2.1.  
 GN Caenorhabditis elegans.  
 OS Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 OC Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J., THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT T., WOHLDSAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;  
 RA WAMSLEY P., BRADSHAW H.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U80033; AAC48199.1; -.  
 SQ SEQUENCE 1847 AA; 199637 MW; B0A25E0F CRC32;

Query Match 60.3%; Score 35; DB 5; Length 1847;  
 Best Local Similarity 58.3%; Pred. No. 3.2e+02;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRILLRY 12  
 DB 19 YRLNVPVLLPY 30

RESULT 10  
 Q85331 ID Q85331 PRELIMINARY; PRT; 179 AA.  
 AC Q85331;  
 DT 01-NOV-1996 (TREMREL. 01, Created)  
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
 DE RNA POLYMERASE 22 KD SUBUNIT (FRAGMENT).  
 OS Vaccinia virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89094998.  
 RA THOMPSON C.L., HOODA-DHINGRA U., CONDIT R.C.;  
 RT "Fine structure mapping of five temperature-sensitive mutants in the 22-and 147-kilodalton subunits of vaccinia virus DNA-dependent RNA polymerase";  
 RL J. Virol. 63:705-713(1989).  
 DR EMBL: M24216; AAA48277.1; -.  
 FT NON\_TER 179 179  
 SQ SEQUENCE 179 AA; 20757 MW; 7E103F8D CRC32;

Query Match 58.6%; Score 34; DB 12; Length 179;  
 Best Local Similarity 50.0%; Pred. No. 54;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRILLRY 12  
 DB 24 YAVINRNVLLRY 35

RESULT 11  
 Q85383 ID Q85383 PRELIMINARY; PRT; 185 AA.  
 AC Q85383;  
 DT 01-NOV-1996 (TREMREL. 01, Created)  
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
 DE HOMOLOG OF VACCINIA VIRUS CDS J4R.  
 GN L4R.  
 OS Variola virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BANGLADESH-1975;  
 RX MEDLINE: 94089747.  
 RA MASSUNG R.F., ESPOSITO J.J., LIU L.I., QI J., UTTERBACK T.R., KNIGHT J.C., AUBIN L., YURAN T.E., PARSONS J.M., LOPAREV V.N.;  
 RT "Potential virulence determinants in terminal regions of variola smallpox virus genome";  
 RL Nature 366:748-751(1993).  
 DR EMBL: L22579; AAA60829.1; -.  
 SQ SEQUENCE 185 AA; 21355 MW; 3B9F48DE CRC32;



Query Match 58.6%; Score 34; DB 12; Length 185;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YELLIRILLRY 12  
| : : | : |||  
DB 24 YAVINRILLRY 35

RESULT 12  
P97213 PRELIMINARY; PRT; 313 AA.  
ID P97213;  
AC P97213;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-MAY-1997 (TREMBlrel. 03, Last annotation update)  
DE CDU2, CDU1, TCDD, TCDB, TCDE, TCDC, CDD1, CDD2, CDD3, AND CDD4  
DE GENES.  
GN CDU2.  
OS Clostridium difficile.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VP110463;  
RA VON EICHEL-STREIBER C.; to the EMBL/GenBank/DBJ databases.  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X92982; CAA63559.1; -.  
DR EMBL; X92982; CAA63558.1; -.  
SQ SEQUENCE 313 AA; 33380 MW; 4F20347A CRC32;

Query Match 58.6%; Score 34; DB 2; Length 313;  
Best Local Similarity 41.7%; Pred. No. 92;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 YELLIRILLRY 12  
| : : | : |||  
DB 259 YRVVRLRLSIKY 270

RESULT 13  
Q9XB05 PRELIMINARY; PRT; 325 AA.  
ID Q9XB05  
AC Q9XB05;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE MEMBRANE ASSOCIATED PROTEIN.  
GN TAD.  
OS Myxococcus xanthus.  
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteriia;  
OC Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ER-15;  
RA PAITAN Y., ORR E., RON E.Z., ROSENBERG E.;  
RT "Genetic and functional analysis of genes required for the post-modification of the polyketide antibiotic TA of Myxococcus xanthus.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ132503; CAB46503.1; -.  
SQ SEQUENCE 325 AA; 35985 MW; 4CC64E85 CRC32;

Query Match 58.6%; Score 34; DB 2; Length 325;  
Best Local Similarity 58.3%; Pred. No. 95;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 YELLIRILLRY 12  
| : : | : |||  
DB 202 YRLTVDRFLRY 213

RESULT 14  
O00326 PRELIMINARY; PRT; 393 AA.  
ID O00326  
AC O00326;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE PROTAGLANDIN EP3 RECEPTOR SUBTYPE ISOFORM.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-UTERUS;  
RX MEDLINE; 97230456.  
RA KOTANI M., TANAKA I., OGAWA Y., USUI T., TAMURA N., MORI K.,  
RA NARUMIYA S., YOSHIMI T., NAKAO K.;  
RT "Structural organization of the human prostaglandin EP3 receptor subtype gene (PTGER3).";  
RL Genomics 40:425-434(1997).  
DR EMBL; D86098; BAA19559.1; -.  
DR PFM; PF00001; 7cml.1; -.  
DR PRINTS; PR00428; PROTAGLNDNR.  
DR PRINTS; PR00582; PRSTNOIDE3P3.  
DR PRINTS; PR00584; PRSTNOIDE32R.  
SQ SEQUENCE 393 AA; 43683 MW; 4EC7995B CRC32;

Query Match 58.6%; Score 34; DB 4; Length 393;  
Best Local Similarity 77.8%; Pred. No. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11  
| : : | : |||  
DB 347 LLIRILLR 355

RESULT 15  
O00325 PRELIMINARY; PRT; 402 AA.  
ID O00325  
AC O00325;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE PROTAGLANDIN EP3 RECEPTOR SUBTYPE ISOFORM.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-UTERUS;  
RX MEDLINE; 97230456.  
RA KOTANI M., TANAKA I., OGAWA Y., USUI T., TAMURA N., MORI K.,  
RA NARUMIYA S., YOSHIMI T., NAKAO K.;  
RT "Structural organization of the human prostaglandin EP3 receptor subtype gene (PTGER3).";  
RL Genomics 40:425-434(1997).  
DR EMBL; D86097; BAA19558.1; -.  
DR PFM; PF00001; 7cml.1; -.  
DR PRINTS; PR00428; PROTAGLNDNR.  
DR PRINTS; PR00582; PRSTNOIDE3P3.  
DR PRINTS; PR00584; PRSTNOIDE32R.  
SQ SEQUENCE 402 AA; 44937 MW; 6A6BE279 CRC32;

Query Match 58.6%; Score 34; DB 4; Length 402;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLIRILLR 11  
| : : | : |||  
DB 347 LLIRILLR 355

Wed Feb 9 08:47:29 2000

us-08-653-294-16.rspt

Page 6

Search completed: February 8, 2000, 13:17:44  
Job time: 32493 sec

OM of: US-08-653-294-16 to: GenEmbl.\* out\_format : pfs

Date: Feb 8, 2000 4:41 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frameat.p2n.model -DEV=xlp  
-O=/cgl1/USPTO.spool/US08653294/runat\_04022000\_160701\_15779/app\_query.fasta.1  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500  
-FGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cd1  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-16

Query length: 12

Database: GenEmbl.\*

Database sequences: 821193

Database length: -1518192014

Search time (sec): 11370.480000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_in1:CEFI4B6	+	45.00	112.32	311.20	29283	! Z81502 Caenorhabditis elegans
gb_btg1:CEFA7H10	+	45.00	92.89	3.8e+03	296589	! Z95311 Caenorhabditis elegans
gb_btg2:HSJ0655C5	+	44.00	89.02	6.2e+03	306959	! AL121956 Homo sapiens chrom
gb_btg5:AC014422	+	43.00	99.17	1.7e+03	59865	! AC014422 Drosophila melanog
gb_bt1:AB007931	-	42.00	113.43	269.92	7150	! AB007931 Homo sapiens mRNA for
gb_bt3:AC013482	+	42.00	92.87	3.8e+03	82875	! AC013482 Genomic sequence for
gb_bt3:AC005365	+	42.00	92.54	3.9e+03	86130	! AC005365 Homo sapiens chromos
gb_bt1:HS1141E20	-	42.00	91.47	4.5e+03	97906	! AL109912 Homo sapiens chromo
gb_bt3:HSJ111206	-	42.00	88.75	6.4e+03	135305	! AL080317 Human DNA sequence
gb_bt7:AC013772	-	42.00	88.30	6.8e+03	142796	! AC013772 Homo sapiens clone
gb_bt3:HSU080134	+	41.00	134.59	17.88	375	! U08134 Human immunoglobulin hea
gb_bt2:CN501C8A	-	41.00	129.98	32.32	650	! AL115310 Botrytis cinerea strai
gb_bt2:CN501C8B	-	41.00	129.24	35.54	710	! AL115311 Botrytis cinerea strai
gb_bt1:UEU131275	-	41.00	127.58	43.96	865	! AJ131275 uncultured euryarchae
gb_bt1:AF142982	-	41.00	127.15	46.42	910	! AF142982 uncultured archaeon B
gb_bt1:AF015992	-	41.00	125.99	53.93	1046	! AF015992 uncultured archaeon 2
gb_bt1:AF015971	-	41.00	125.98	53.98	1047	! AF015971 uncultured archaeon 2
gb_bt1:AF015978	-	41.00	125.98	53.98	1047	! AF015978 uncultured archaeon 2
gb_bt1:AB013391	+	41.00	93.25	3.6e+03	51737	! AB013391 Arabidopsis thaliana
gb_bt6:AC016132	+	41.00	91.14	4.7e+03	66465	! AC016132 Drosophila melanog
gb_bt6:AC015445	+	41.00	87.36	7.6e+03	104365	! AC015445 Arabidopsis thalian
gb_bt3:HS3393P12	+	41.00	87.34	7.7e+03	104597	! AL022578 Homo sapiens DNA se
gb_bt3:AF037352	+	41.00	87.25	7.8e+03	105726	! AF037352 Mus musculus T cell
gb_bt4:AC008464	+	41.00	85.49	9.7e+03	130435	! AC008464 Homo sapiens chrom
gb_bt4:AC007456	+	41.00	79.53	2.1e+04	265232	! AC007456 Homo sapiens chrom
gb_bt3:AC009602	+	40.00	91.83	4.3e+03	40000	! AC009602 Leishmania major chr
gb_bt3:AC008413	-	40.00	87.90	7.1e+03	63934	! AC008413 Homo sapiens chromos
gb_bt2:AC002453	-	40.00	87.10	7.9e+03	70319	! AC002453 Human BAC clone GS44
gb_bt2:AC006300	+	40.00	85.14	1.0e+04	88770	! AC006300 Arabidopsis thaliana
gb_bt2:AC004368	-	40.00	85.04	1.0e+04	89885	! AC004368 Drosophila melanog
gb_bt4:AC011180	-	40.00	80.99	1.7e+04	145531	! AC011180 Homo sapiens clone
gb_bt2:AF165139	+	40.00	80.43	1.8e+04	155664	! AF165139 Homo sapiens chrom
gb_bt7:AC017155	+	40.00	79.99	2.0e+04	163954	! AC017155 Drosophila melanog
gb_bt4:AC006377	+	40.00	79.38	2.1e+04	176352	! AC006377 Homo sapiens chrom
gb_bt4:AC006377	+	40.00	79.38	2.1e+04	176463	! AC006377 Homo sapiens clone
gb_bt6:AC016259	+	40.00	79.22	2.2e+04	179656	! AC016259 Homo sapiens chrom
gb_bt1:CN501D7M	+	40.00	78.69	2.3e+04	191442	! AL132800 Homo sapiens chrom
gb_bt1:AF142981	-	39.00	122.34	86.07	689	! AF142981 uncultured archaeon B
gb_bt1:UEU131254	-	39.00	120.44	109.81	864	! AJ131254 uncultured euryarchae
gb_bt1:UEU131274	-	39.00	120.44	109.81	864	! AJ131274 uncultured euryarchae
gb_bt1:UEU131272	-	39.00	120.43	109.95	865	! AJ131272 uncultured euryarchae
gb_bt1:UEU131265	-	39.00	120.42	110.08	866	! AJ131265 uncultured euryarchae

gb\_ba2:AF134389 - 39.00 120.09 114.88 901 ! AF134389 Uncultured archaeon  
gb\_ba2:AF134390 - 39.00 120.08 115.02 902 ! AF134390 Uncultured archaeon  
gb\_ba1:UARI33622 - 39.00 117.04 169.89 1296 ! AJ133622 Uncultured archaeon

seq\_name: gb\_in1:CEFI4B6

seq\_documentation\_block:

LOCUS CEFI4B6 29283 bp DNA INV 02-SEP-1999  
DEFINITION Caenorhabditis elegans cosmid F14B6, complete sequence.

ACCESSION Z81502

VERSION Z81502.1 GI:2814053

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 29283)

Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,  
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,  
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,  
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,  
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,  
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,  
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra.A.,  
Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,  
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,  
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,  
Wilkinson-Sproat,J. and Wohlman,P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans

JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE 94150718

REFERENCE 2 (bases 1 to 29283)

AUTHORS White,S.

TITLE Direct Submission

JOURNAL Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwenematode.wustl.edu

On Jan 28, 1998 this sequence version replaced gi:1665918.  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFinder (P. Green, U. Washington), and other  
available information.  
For a graphical representation of this sequence and its analysis  
see:  
http://webcbs.sanger.ac.uk/cgi-  
bin/display?db=wormacsc&sequence=subject=F14B6  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
IMPORTANT: This sequence is not the entire insert of clone F14B6.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone F41D3 is at 29177 in this sequence. The  
true right end of clone C15C6 is at 19321 in this sequence. The  
start of this sequence (1..109) overlaps with the end of sequence  
Z81034.  
The end of this sequence (29177..29283) overlaps with the start of  
sequence Z81537.

FEATURES

source

location/Qualifiers  
1. .29283  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="I"  
/clone="F14B6"  
complement(947..2169)  
/gene="F14B6.1"  
complement(join(947..1390,1835..2063,2114..2169))

```

/gene="F14B6.1"
/note="cDNA EST yk242e8.5 comes from this gene; cDNA EST
yk242e8.3 comes from this gene; cDNA EST yk414b10.5 comes
from this gene; cDNA EST yk477b2.3 comes from this gene;
cDNA EST yk477b2.5 comes from this gene; cDNA EST
yk608e9.3 comes from this gene"
/codon_start=1
/protein_id="CAB04105.1"
/db_xref="GI:3875879"
/db_xref="SPTREMBL:O62161"
/translation="MFSVEWILIFFGILNWKFTQPTPEPASRFFTRVCOOLRETV
LIEIDSSGRVQFVSEEGSKPYFPHVGCACQWMDVYGEQPHRVEEDQDTE
GVSEIEIDSDFDQITFKPTTAPLSPETENFIEDKMKHFQATIDINFLGKGT
DMGIYNLRISQITGVOLKQVSSGSKSMKEGILSKLQKEFEIEGLNGEISEAE
ENESDIDSEFEVENLELDSDFE"
complement(2938..3345)
/gene="F14B6.7"
complement(join(2938..3065,3115..3212,3260..3345))
/gene="F14B6.7"
/codon_start=1
/protein_id="CAB54218.1"
/db_xref="GI:5824443"
/translation="MGICVPRWSLFLVHLIVDFLVLGVFTVSLIAIMSSIMYFLID
EMTSPLFLFLVAFVLDIVLILYAIWGIAYRCCCLVDSIKETSINYSVSSNGEA
V"
complement(4016..6657)
/gene="F14B6.2"
complement(join(4016..4423,4517..5309,5485..5820,
6093..6275,6563..6657))
/gene="F14B6.2"
/codon_start=1
/protein_id="CAB04106.1"
/db_xref="GI:3875880"
/db_xref="SPTREMBL:O62162"
/translation="MSTSSVLPPTPIGKQVVIDLKAGAILRENARKSVLFGCPLSV
VLIVLILKLCISWIRHEHETDFRCAAWNGQGTDMYRAKRSIAFEKSDPRKFS
KIPDDDDNSDNDVINKDKKKKPLYGSTSPANDIISDAPKIDFLDGFPLPNFQ
RIDDSHPAVPPYDPPDPLMTWLTNTQIVOLSGEETGVWVMKTIIRNVEKVG
DEOFTDIANAPRIVSVKIKSRGGEHEEHFDLLKELKDEEKKKKKKGKGR
LKAKKLTAEKLLNNTSVLPDTANVRKPKGLKSKKLKPKSDERTVTKQPR
HLVGSSLSPTTPRPSKSWMGVLDQDPNFSNOGQINGYPMQAKSRKPNATSPS
EKDEEEKLSESTRMMWMDLKYITFLAPKTEQTRQVEEGVKEVIOEAHERSQ
PKLHTEFASSTPTPPAFHFIPPSSEAPYYVEINDADTETVYVSSVTPMYRPFEN
APQHIFESASPTPTPAFHFIPPSSEAPYYVEINDADTETVYVSSVTPMYRPFEN
MGLHSTKSYDPLHLEVSFSEFCVGLLCHFEVDP"
complement(9869..9940)
/gene="F14B6.t2"
/note="CAT Met M-tRNA; predicted using trnscan-se-1.11;
preliminary prediction; similar to trna-met"
complement(9869..9940)
/gene="F14B6.t2"
13752..13823
/gene="F14B6.t1"
/note="CAT Met M-tRNA; predicted using trnscan-se-1.11;
preliminary prediction; similar to trna-met"
13752..13823
/gene="F14B6.t1"
13024..16695
/gene="F14B6.3"
join(15024..15079,15134..15329,15847..16695)
/gene="F14B6.3"
/note="predicted using Genefinder; cDNA EST yk329e12.5
comes from this gene; cDNA EST EMBL:C08202 comes from this
gene; cDNA EST EMBL:C09873 comes from this gene; cDNA EST
yk198f11.3 comes from this gene; cDNA EST yk198f11.5 comes
from this gene; cDNA EST yk329e12.3 comes from this gene;
cDNA EST yk338a7.3 comes from this gene; cDNA EST
yk338a7.5 comes from this gene; cDNA EST yk404g5.3 comes
from this gene; cDNA EST yk404g5.5 comes from this gene;
cDNA EST yk415a6.3 comes from this gene; cDNA EST
yk415a6.5 comes from this gene; cDNA EST yk498b10.3 comes
from this gene; cDNA EST yk509c5.3 comes from this gene;
cDNA EST yk573a12.3 comes from this gene"

```

```

/codon_start=1
/protein_id="CAB04104.1"
/db_xref="GI:3875878"
/db_xref="SPTREMBL:O62160"
/translation="MLSFLQSFRLSLGYEDYFKFIAPSSPRFFTKYCNETOKKTIIVI
DSNNIIRPFVEDSENEREPLEVAHLGSCSECRWNRMYGENIAENNETPDEDETDYE
SETSDSELQSDPNPNSALNNCIENFDQLVDEVRKDFHRSQQLRLVEECIGNM
NSFLQENLTVDDISTGACALLNETAKLQMFSSSGISGRKSESEGTNKDILKIFGEL
RIGFQNEEDDDDESVEGEEDSDDEDLDDQEGESSIEPRLSETFKKSARELSLS
KEAPVKDLVIYKPRDEHEDLESENERMSDEAKLVEDKLIIESEDESDIIGRDE
ELDDSNKKEEELELOVQDESDDEQFFSL"
17155..19537
/gene="F14B6.6"
join(17155..17218,17267..17455,17898..18232,18389..18591,
19031..19178,19319..19537)
/gene="F14B6.6"
/note="cDNA EST yk549c2.3 comes from this gene"
/codon_start=1
/protein_id="CAB04109.2"
/db_xref="GI:5824444"
/db_xref="SPTREMBL:O62165"
/translation="MKARKILITVAIILASLALRYYSORTMEFLIEEDOMCLOREWIE
SEFCQSEILNUNFKFSKLFKKSINFINNIYSKFTISSQINSTADYGNKQISF
SDVOSTFKWIHIPKISGSEILLIVLRPDDFSRRNVRHRTWMSQENEIKYLFILGL
ANMDGKIYVMAEAKLFGDIWVTSMEDYRSKLSFKTILLLFGVSKVPSAQLIGKID
GDLVFPNLFSTIKNENMINVTSASVYKIAEAGVPIVSNCKRFLKYNFRNSLI
SFGTRVAFYLAGPFLVTRSAALRLLEASKHRNFHKIEDTLITGLVADDDVSRVQL
HRINLQEGKGTDLVFAWHSPLNDPEYKDLIYKTMSSQQFKEKRLQELNVI"
20434..22684
/gene="F14B6.4"
join(20434..20524,21137..21250,21298..21659,21969..22168,
22218..22344,22393..22482,22538..22684)
/gene="F14B6.4"
/codon_start=1
/protein_id="CAB04107.1"
/db_xref="GI:3875881"
/db_xref="SPTREMBL:O62163"
/translation="MEISLQKKLFLFLFANASSVSGSGSDYVLLKKSISGLNAYQID
KNMKEHEHKLALTNCKMGRINSSVPTKDFGSSFTISFADIKTHSWLYLPKFO
NSMSEILMIVASTDTSFARNVLRKTMNPENSEILKDGEMKALFLVGTWDGDSRM
KRYMEARIYGMVYVDLKYTEELFSLTLLYGTSKASEFKLGKIDEDIMFFP
DLIPLQENLNDPSSSIYGMFAEGGYVTRKREHRWFVPDSTYGCMPFPYTGGLF
KVLTPLEAKRLINATKRIPIEDIALINGLANDCKIPRIHLPIYGDHHSLLDEDT
RGVLAWHTGKNDQQLDYFYRKLSESRLLKKKILDEKNKHLEY"
23603..26251
/gene="F14B6.5"
join(23603..23727,23784..24048,24104..24209,24287..24373,
24425..24635,24799..24965,25030..25343,25388..25514,
25566..25680,25729..25933,25988..26251)
/gene="F14B6.5"
/codon_start=1
/protein_id="CAB04108.1"
/db_xref="GI:3875882"
/db_xref="SPTREMBL:O62164"
/translation="MSVAAKPKRLDLQIRGAILSLVGLGHFFYPALFPNGYLGVDQFF
ETNOKSAMRALLFVSNAKPSQDDYFTMLARAVDIFHTWLSLSEVOFYFLVPTFLF
VLSGFLMCLLKRAENETPTCLITLYSKRKRILPLYLLILLMSISLYTFPDTSTI
ETNOKSAMRALLFVSNAKPSQDDYFTMLARAVDIFHTWLSLSEVOFYFLVPTFLF
AIKLSGLSYAFFFTPTNSFNSFARINQOFLGMIVFLLSHSSKSPKAKIEEC
NKLDEENNTOKAKPSILHLISYFFLISILICIPIQNTDFVPLVLTATGALM
LISDSNLLSNKILTYIGDISYSLIHWPIYAKWLAFGDGVNLLILSLSLSLAI
IVFFEFEKWKYKLSSTIGLILVFLFLNVLINKDELIPDRENNNNHSSMNTDDAI

```

```

alignment_scores:
  Quality: 45.00      Length: 12
  Ratio: 4.091      Gaps: 0
  Percent Similarity: 91.667      Percent Identity: 66.667

alignment_block:
  US-08-653-294-16 x CEF14B6 ..
  Align seq 1/1 to: CEF14B6 from: 1 to: 29283
  1 TyrArgLeuLeulleArgArgIleLeuLeuArgTyr 12

```

```
|||||
3030 TACCGACTACTCGCGTAGAGTATTAAACAATAT 3065
seq_name: gb_htg1:CEY47H10

seq_documentation_block:
LOCUS CEY47H10 296589 bp DNA HTG 07-SEP-1999
DEFINITION Caenorhabditis elegans chromosome I clone Y47H10, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION 295311
VERSION 295311.10 GI:5725202
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 296589)
AUTHORS McKay, K.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RO, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jess@sanger.ac.uk or rwenematode.wustl.edu
COMMENT On Aug 10, 1999 this sequence version replaced gi:4938508.
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1..296589
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="I"
/clone="Y47H10"
BASE COUNT 92991 a 55009 c 53606 g 92582 t 2401 others
ORIGIN

alignment_scores:
Quality: 45.00 Length: 12
Ratio: 4.091 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 66.667

alignment_block:
US-08-653-294-16 x CEY47H10 ...
Align seg 1/1 to: CEY47H10 from: 1 to: 296589
1 TyrArgLeuLeuLeuArgArgilleLeuLeuArgTyr 12
|||||
188134 TACCGACTACTCGCGTAGAGTATTAAACAATAT 188169
seq_name: gb_htg2:HSDJ655C5

seq_documentation_block:
LOCUS HSDJ655C5 306999 bp DNA HTG 26-NOV-1999
DEFINITION Homo sapiens chromosome 6 clone RP4-655C5, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL21956
VERSION AL21956.2 GI:6469398
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 306999)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 306999)

Sims, S.
Direct Submission
Submitted (26-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Nov 27, 1999 this sequence version replaced gi:6066017.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00008 Length: 1036bp
Contig_ID: 00019 Length: 9963bp
Contig_ID: 00031 Length: 1518bp
Contig_ID: 00039 Length: 1498bp
Contig_ID: 00059 Length: 1840bp
Contig_ID: 00062 Length: 1083bp
Contig_ID: 00088 Length: 1408bp
Contig_ID: 00092 Length: 1477bp
Contig_ID: 00112 Length: 3742bp
Contig_ID: 00142 Length: 1888bp
Contig_ID: 00150 Length: 3371bp
Contig_ID: 00179 Length: 1299bp
Contig_ID: 00183 Length: 3722bp
Contig_ID: 00192 Length: 1260bp
Contig_ID: 00211 Length: 2606bp
Contig_ID: 00242 Length: 1648bp
Contig_ID: 00248 Length: 4175bp
Contig_ID: 00272 Length: 1781bp
Contig_ID: 00323 Length: 10365bp
Contig_ID: 00355 Length: 1790bp
Contig_ID: 00387 Length: 4242bp
Contig_ID: 00394 Length: 6397bp
Contig_ID: 00404 Length: 1731bp
Contig_ID: 00421 Length: 1266bp
Contig_ID: 00476 Length: 1153bp
Contig_ID: 00481 Length: 980bp
Contig_ID: 00528 Length: 2380bp
Contig_ID: 00555 Length: 2876bp
Contig_ID: 00567 Length: 1458bp
Contig_ID: 00603 Length: 2634bp
Contig_ID: 00605 Length: 1049bp
Contig_ID: 00638 Length: 1196bp
Contig_ID: 00679 Length: 3286bp
Contig_ID: 00695 Length: 4879bp
Contig_ID: 00709 Length: 3049bp
Contig_ID: 00722 Length: 5922bp
Contig_ID: 00725 Length: 1901bp
Contig_ID: 00750 Length: 1420bp
Contig_ID: 00812 Length: 1513bp
Contig_ID: 00892 Length: 1333bp
Contig_ID: 00914 Length: 14732bp
Contig_ID: 00940 Length: 1506bp
Contig_ID: 00964 Length: 1687bp
Contig_ID: 00973 Length: 1044bp
Contig_ID: 00979 Length: 1055bp
Contig_ID: 01017 Length: 3425bp
Contig_ID: 01057 Length: 3233bp
Contig_ID: 01085 Length: 1282bp
Contig_ID: 01125 Length: 3177bp
Contig_ID: 01137 Length: 3871bp
Contig_ID: 01155 Length: 1111bp
Contig_ID: 01212 Length: 1153bp
Contig_ID: 01219 Length: 5488bp
Contig_ID: 01220 Length: 1788bp
Contig_ID: 01238 Length: 3851bp
Contig_ID: 01243 Length: 4555bp
Contig_ID: 01246 Length: 1021bp
Contig_ID: 01256 Length: 1097bp
Contig_ID: 01288 Length: 3286bp
Contig_ID: 01296 Length: 5117bp
Contig_ID: 01338 Length: 2253bp
```

ADAMS, M. and Venter, J.C.  
Direct Submission  
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
This sequence was identified as CDM:10210882 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1..59865  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 16683 a 13092 c 13361 g 16729 t  
ORIGIN

alignment\_scores:  
Quality: 43.00 Length: 12  
Ratio: 3.909 Gaps: 0  
Percent Similarity: 91.667 Percent Identity: 66.667

alignment\_block:  
US-08-653-294-16 x AC014422/rev ..  
Align seg 1/1 to reverse of: AC014422 from: 1 to: 59865  
1 TyArgLeuLeuLeuArgArgIleLeuLeuArgTyr 12  
|||||  
|||||  
49675 TATCGACTATTGTCATCGATCTTGTTCGATAC 49640

seq\_name: gb\_pri:AB007931  
seq\_documentation\_block:  
LOCUS AB007931 7150 bp mRNA PRI 13-AUG-1998  
DEFINITION Homo sapiens mRNA for KIAA0462 protein, partial cds.  
ACCESSION AB007931  
VERSION AB007931.1 GI:3413885  
KEYWORDS KIAA0462 protein.  
SOURCE Homo sapiens male brain cDNA to mRNA, clone\_lib:pBluescriptII SK  
plus clone:HG0891.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 7150)  
AUTHORS Ohara, O.  
TITLE Direct Submission  
JOURNAL Submitted (08-OCT-1997) to the DBJ/EMBL/GenBank databases. Osamu  
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,  
Fax:+81-438-52-3914)  
2 (slices)  
SEKI, N., OHIRA, M., NAGASE, T., ISHIKAWA, K., MIYAJIMA, N.,  
NAKAJIMA, D., NOMURA, N. and OHARA, O.  
TITLE Characterization of cDNA clones in size-fractionated cDNA libraries  
from human brain  
JOURNAL DNA Res. 4 (5), 345-349 (1997)  
MEDLINE 98116602  
FEATURES  
source  
Location/Qualifiers  
1..7150  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="HG0891"  
/clone\_lib="pBluescriptII SK plus"  
/sex="male"  
/tissue\_type="brain"  
1..6832  
/gene="KIAA0462"  
/gene="KIAA0462"

gene  
CDS

```

/codon_start=2
/product="KIAA0462 protein"
/protein_id="BAA32307.1"
/db_xref="GI:3413886"
/translation="EHSVSGRSSAYGATAGHAGPAGSGSVSSSTGALSTTTTGHQEGG
SGEGEGEGEDVTHNKLHVRMLLRLTLPQLRNVGGVRAIPYQVIMLMTTD
LDGEDEKDGALDNLISQIAELGMDKDDVSKNERSALNEVHLVWMLLSVFSRTK
SGKSSSCESSSSATAAALLSGAVDYCLHYLKSLEYKWSQQNDPEFVRSQLL
KPHSTSPDMPSPFELQYVKGHAADVFEAYTOLLTEMLRLPYOIKKIDTNSRIPL
VPFDSWFYFLSEVYIQQOTPEVRORVKLLFLICGSKERYRLDRHTLDHVRGKIK
KLLEQGIIFRASVYVTSAGSALOYDILLSMEHLKACAEIAAQTINWQKFCIKDS
VLYFLQVSLVDGSPVQLQLSCALSKYLAALAASSSSSSSSAPVAASG
QATTSKSSSKKEKEKEKDEGTSGSQEDQLCTALYNOLNFKADETLQYLRFC
LLENSNSVQWACHLTHIYRNSKSOQELLDLMSWPELPAFGKRAQFVLLIG
YFSCLTPQTEKKLEYQKAVEILRTQNLHTNHPNSNIYNTSLGVFFQGYLSESD
CLVGNPPEPCYIKLSIKVDTRYTTQOVVKGISHTISKVTYKIDGLKRTKWRT
INLYNRTQVAIVELKNKPARWHKAKKVLTPGTEVKIDLPPIVASNLMEFADF
YENYQASTELQPCRSASVPANFCVCGNGENYQCHKRSINDEDFPLCNACGF
XKARFDFMLAKPCCCAVDPIENEEDRRKAVSNITLDRADRYHQLMGRPOLENL
LCKYNEAPEKPDQDSRTAGGISSTASVNYIILQAOEYCGCKNSFDELSKTIQKV
FASRKELEYPDQREAAKSSRTSVQPTASQYRALSVLGCGHTSTKCYGCASAV
TECHITLLRALNPALRHILVSQGLIRELFDYNLRRGAAMREEVQLMKLTRDNP
EATQOMDLIIGKYSTALKGHWNPDLASSIQLYEMLLITDSISKEDSCWELRCLALS
LFLMAVNIKPVVVENTILMCLRLILQKLIAPPATSKNNDVPVEALTYKPYCNEIH
AQOLWLRPKPKASYDAWKCLPILRGIDGNGKAPSKSELRLHYLTERKYVWRKQFLSR
RGKRTSLDLKLGHNRLRVLPATQAARQAQCTIVEALATIPSRKQVLDLITSY
LDLSIAGECAAEYALYOKLITSAHKVKYLAARGVLPYVGNLTIKELIARLLAEAT
LSTDQOYALKSLTGLLSSEFVESIKRHKFSRLVGLVNGLYLCKRLVQVTRKLLID
ETQDMLEMLDMTGTETSEKAPMAVCITAKYNNDDITPTVIFERLCSTIIPBE
NEVFTEVLEKQDQEDFQGRMPGNPYISNEPGIPLMRDINKKICQCDVALLE
DDSGMELLVNNKILSLDPAEYVYKVMCTTNEGEPRIYVRMGLGDATEEFIESL
DSTDEDEEEVYKMAVQAQCGLECMNLRLAGIDFKQGRHLTLVLLKLSYCVK
YKVRQOLVLEMTLNVMLGTNLALVAQESKDSGAAVAEOVLSIMEIILDESNA
EPLSEKGNLLGDKQVLMDDOINSTRVSRNPVYGLLRIPIPLSGEVEKQOI
LYVERKFCNFDYDEHSDGDKVFLCFCFKIAGIANNNGHQLKDLILQKLTQNA
LDYMKHIPSANKNLDLWKKFLSRPALPFIILRLGLAIOHPGTQVLTGDTSPNHL
KLQVSDSEGITLAEHLREHPDVKKIDAAARETRAERKRMAMARQKALGFL
MTNTEGQVVTALLKOMEELIEEPLGTCRCREGYKQPTKVLGYITFTKRVALE
EMENKPKQGYSTVSHFNIVYDCHLAAVLARGREWSAALONCNGLLPVM
GPHVPESAQCTALRHNTYLOEGTGOREPTVOLNHDIKLLEFRAMEQSFADLPGV
GRESNIHLIPIYIHLVLTNPATSRKKNLQGLFQEPKKEVSAFEDVGPYFT
VLALHILPPQWRATRVETLRLLTVSQARAPGGATRLTDRAKDYSAIRSSLLFW
ALVDLIYNKQKVPNTSGEGWSCSLAEYIRHNDMPYIEAADKALKTQBEFMPVTF
SEFLDVAGLLSEITDPESFLKDLNLSVP"
BASE COUNT 1797 a 1916 c 1861 g 1576 t
ORIGIN

alignment_scores:
  Quality: 42.00      Length: 12
  Ratio: 3.818
Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:
US-08-653-294-16 x AB007931/rev ..

Align seg 1/1 to reverse of: AB007931 from: 1 to: 7150

1 TyrArgLeuLeuLeuArgArgIleLeuLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||:|||||:
2058 TACAGGTTGATGTCGCGACCATCTTGTCCGTTTC 2023

seq_name: gb_pl3:AC013482

seq_documentation_block:
LOCUS AC013482 82875 bp DNA PLN 09-DEC-1999
DEFINITION Genomic sequence for Arabidopsis thaliana BAC T26F17 from
chromosome 1, complete sequence.
ACCESSION AC013482
VERSION AC013482.2 GI:6478885
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 82875)
Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
Khan,S., Kim,C., Altati,H., Bei,Q., Chin,C., Chioiu,J., Choi,E.,
Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G.,
Federspiel,N.A., Theologis,A. and Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC T26F17 from
chromosome 1
Unpublished
2 (bases 1 to 82875)
Ecker,J.R.
Direct Submission
Submitted (13-NOV-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 82875)
Ecker,J.R.
Direct Submission
Submitted (30-NOV-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 82875)
Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
Khan,S., Kim,C., Altati,H., Bei,Q., Chin,C., Chioiu,J., Choi,E.,
Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
Federspiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (09-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Nov 30, 1999 this sequence version replaced gi:6403470.
Location/Qualifiers
  1. 82875
    /organism="Arabidopsis thaliana"
    /db_xref="taxon:3702"
    /chromosome="1"
    /clone="n26F17"
    /join(140..398,676..953,1145..1474)
    /note="unknown; similar to EST gb|H74506"
    /codon_start=1
    /evidence=not_experimental
    /product="T26F17.1"
    /protein_id="AAE16527.1"
    /db_xref="GI:6552728"
    /translation="MARGISGGKNGAGQFFLATLLWLVSVVFETAFNRTTELL
WVICGCGFTQLMNVVRSWISFTSIVSLLHSITSAVVFILLNQCLAKGLDE
MFDSELVGGTWKWAYPALFCGCFAYQDWMLOYRSLGSLPVLVHLLVLCVFT
LALYRNVTNYLITLICEMHSITFLHVRKLRMAGIRDSNTALVKLEWLNWTFVFA
RCIPHLITIKLIKDAHKFGKGVLPALSGMAGMNLNVGLDILDFHAFRRSRHR
NQENGLDHGE"
    complement(Join(2361..2747,2834..4239,4324..4420))
    /note="hypothetical protein"
    /codon_start=1
    /evidence=not_experimental
    /product="T26F17.2"
    /protein_id="AAE16547.1"
    /db_xref="GI:6552748"
    /translation="MMTGTTILEPVMDSKDELVKQAKVAEDAVAGWEKAENEVVEL
KOKLEDAADKNIVLEDVSHLDGALKECVROLRFQDEQKNIOAAVTSTKELHSAN
TGLEKVELIQKEAAKSENMMRLREFLTQREDEIVMERDLSQAAETASKQHLH
IINKLAKAECKRLIKARTSSSSNSGSDSHGGERVEGSDSCSWASSAFISE
LQIKNEKGNRSLOQTTSSTIDMLDMEFLERLVALPTQAKNKDGYELSLMEK
LEKIOAKDDLEVEKAEKRLSEIAYVVGKVELEDMLKRYEAKELKTSFD
VLKDKYQSRVCFQVDTTKLEKLAQAEKDELSDSEVICCKEAEKRFSLSELEAVVGDKTEM

```





Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 86130)  
 Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,  
 Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,  
 Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,  
 White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,  
 Misra,M. and Deaven,L.  
 Sequencing of Human Chromosome 16  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 86130)  
 Ricke,D.O.  
 Large Scale Sequence Analysis and Annotation with the Sequence  
 Comparison Analysis (SCAN) System  
 JOURNAL  
 Unpublished  
 3 (bases 1 to 86130)  
 Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,  
 Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,  
 Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,  
 White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,  
 Misra,M. and Deaven,L.  
 Direct Submission  
 JOURNAL  
 Submitted (01-AUG-1998) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
 FEATURES  
 source  
 1. 86130  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="79-2A"  
 2000. .2142  
 /note="GRAIL 2 excellent exon, frame 2"  
 2185. .2497  
 /rpt\_family="Alu"  
 7853. 8448  
 /rpt\_family="Alu"  
 9717. .10001  
 /rpt\_family="Alu"  
 10342. .10563  
 /note="(T)22"  
 /rpt\_type=tandem  
 /rpt\_unit=T  
 complement(10546. .10672)  
 /rpt\_family="Alu"  
 11817. .12730  
 /note="GRAIL 2 excellent exon, frame 0"  
 14804. .14902  
 /note="GRAIL 2 excellent exon, frame 2"  
 15869. .16008  
 /note="GRAIL 2 excellent exon, frame 1"  
 16275. .16548  
 /rpt\_family="Alu"  
 17124. .17355  
 /rpt\_family="MER33"  
 complement(18448. .18747)  
 /rpt\_family="Alu"  
 19083. .19373  
 /rpt\_family="Alu"  
 19357. .19380  
 /note="(A)24"  
 /rpt\_type=tandem  
 /rpt\_unit=A  
 20307. .20329  
 /note="(T)23"  
 /rpt\_type=tandem  
 /rpt\_unit=T  
 20836. .21048  
 /note="GRAIL 2 excellent exon, frame 0"  
 21252. .21358  
 /rpt\_family="Alu"  
 complement(22142. .22428)  
 /rpt\_family="Alu"  
 27568. .27837  
 /rpt\_family="Alu"

28059. .28174  
 /note="GRAIL 2 excellent exon, frame 0"  
 28975. .29221  
 /rpt\_family="Alu"  
 29408. .29459  
 /rpt\_family="MIR"  
 30146. .30276  
 /rpt\_family="Alu"  
 30513. .30631  
 /note="GRAIL 2 excellent exon, frame 1"  
 30919. .31186  
 /rpt\_family="Alu"  
 31608. .32152  
 /rpt\_family="Alu"  
 complement(32560. .32851)  
 /rpt\_family="Alu"  
 33192. .33293  
 /rpt\_family="Alu"  
 35811. .35929  
 /note="GRAIL 2 excellent exon, frame 2"  
 36035. .36312  
 /rpt\_family="Alu"  
 complement(36626. .36748)  
 /rpt\_family="L1"  
 complement(37173. .37468)  
 /rpt\_family="Alu"  
 38519. .38818  
 /rpt\_family="Alu"  
 40556. .40673  
 /rpt\_family="MIR"  
 41506. .41532  
 /note="(A)27"  
 /rpt\_type=tandem  
 /rpt\_unit=A  
 45596. .45892  
 /rpt\_family="Alu"  
 46081. .46366  
 /rpt\_family="Alu"  
 46477. .46721  
 /note="GRAIL 2 excellent exon, frame 0"  
 47484. .47760  
 /rpt\_family="Alu"  
 50106. .50312  
 /note="GRAIL 2 excellent exon, frame 2"  
 50512. .50788  
 /rpt\_family="Alu"  
 50789. .51068  
 /rpt\_family="Alu"  
 complement(52275. .52538)  
 /rpt\_family="Alu"  
 complement(53058. .53720)  
 /rpt\_family="L1"  
 54695. .54976  
 /rpt\_family="Alu"  
 55556. .55829  
 /rpt\_family="Alu"  
 56565. .56848  
 /rpt\_family="Alu"  
 complement(57163. .57314)  
 /note="GRAIL 2 excellent exon, frame 1"  
 58926. .59183  
 /rpt\_family="Alu"  
 59490. .60014  
 /rpt\_family="Alu"  
 complement(60362. .60655)  
 /rpt\_family="Alu"  
 60988. .61273  
 /rpt\_family="Alu"  
 61417. .61757  
 /rpt\_family="Alu"  
 complement(62028. .62141)  
 /rpt\_family="Alu"  
 complement(63160. .63454)

ORGANISM	Sequence
<i>Homo sapiens</i>	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominidae; Homo.	

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Dec 10, 1999 this sequence version replaced gi:6425750.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center -----

Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

```

Cancer clone name: 10224
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 128649 bases at least Q40
Consensus quality: 137112 bases at least Q30
Consensus quality: 140758 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 142796; sum-of-contents
Quality coverage: 5.7 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-contents

```

[illegible]

1 2142: contig of 2142 bp in length

*	2143	4834:	contig of 2692 bp in length
*			gap of unknown length
*			gap of unknown length
*	4835	9211:	contig of 4377 bp in length
*			gap of unknown length

25158	66208:	contig of 41051 bp in length
*		gap of unknown length
*		gap of unknown length
*		gap of unknown length
66209	142796:	contig of 76588 bp in length
*		gap of unknown length
*		gap of unknown length
*		gap of unknown length

```
Location/Qualifiers
1. .142796
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="RP11-7024"
```

```
47989 a 24580 c 24540 g 45686 t 1
```

cores: 42.00 Length: 12

4.2:00	Length:	12
4.200	Gaps:	0

Percent Similarity: 83.333  
Percent Identity: 41.200

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

ent\_block:

-653-294-16 x AC013772/rev ..

seg 1/1 to reverse of: AC013772 from: 1

1 TyrArgLeuLeuIleArgArgIleLeuLeuArgTyr 12

221 TATAGCTAGCAGTCAGAAGACATTTACTTAAATAT 13

1000

name: gb\_pr3:HSU80134

KEYWORDS	CDNA library; nitrogen deprivation.
SOURCE	<i>Botryotinia fuckelliana</i> .
ORGANISM	<i>Botryotinia fuckelliana</i> .
REFERENCE	Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes; Leotiales; Sclerotiniaceae; Botryotinia.
AUTHORS	1 (bases 1 to 650)
TITLE	Bitton, F., Lewis, C., Fortini, D., Pradier, J.M. and Brygoo, Y.
JOURNAL	Direct Submission
REFERENCE	Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
AUTHORS	2 (bases 1 to 650)
TITLE	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (01-SEP-1999) Genoscope - Centre National de Sequençage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : <a href="mailto:segref@genoscope.cns.fr">segref@genoscope.cns.fr</a> ) - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

```

FEATURES             Location/Qualifiers
     source            ..850
     /organism="Botryotinia fuckeliana"
     /strain="T4"
     /db_xref="taxon:40559"
     /note="Genoscope sequence ID : W60E121"
193 a                121 c                167 t
BASE COUNT
ORIGIN
alignment_scores:
    Quality: 41.00          Length: 12
    Ratio: 3.417           Gaps: 0
    Percent Similarity: 100.000  Percent Identity: 50.000
alignment_block:
US-08-653-294-16 x CNS01CBA/rev ..
Align seg 1/1 to reverse of: CNS01CBA from: 1 to: 650
1 TyrArgLeuLeuIleArgGIIleLeuLeuArgTyr 12
443 TTTCGGATCTTGTTGTCGAAGATATTACTAAAGTTT 408

```

LOCUS	CNS01CBB	710 bp	mRNA	PLN	02-SEP-1999			
DEFINITION	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation.							
ACCESSION	AL115311							
VERSION	AL115311.1	GI:5829930						
KEYWORDS	cDNA library; nitrogen deprivation.							
SOURCE	Botryotinia fuckeliana.							
ORGANISM	Botryotinia fuckeliana							
	Eukaryota; Fungi; Ascomycota; Euscomycetes; Discomycetes; Leotiales; Sclerotiniaceae; Botryotinia.							
REFERENCE	1 (bases 1 to 710)							
AUTHORS	Bitton, F., Lewis, C., Fortini, D., Pradier, J.M. and Brygoo, Y.							
TITLE	Direct Submission							
JOURNAL	Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France							
REFERENCE	2 (bases 1 to 710)							

REFERENCE	2 (bases 1 to 710)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (01-SEP-1999) Genoscope - Centre National de Sequençage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : <a href="mailto:secref@genoscope.cns.fr">secref@genoscope.cns.fr</a> - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal

alignment_scores:		
Quality:	41.00	Length: 12
Ratio:	3.417	Gaps: 0
Percent Similarity:	100.000	Percent Identity: 50.000

**THIS PAGE BLANK (USPTO)**



76 AGGTACTCTCGAGGAGGATCTCTTAACATT 44

seq\_name: N\_Geneseq\_36:T19381

seq\_documentation\_block:

ID T19381 standard; cDNA to mRNA; 119 BP.

AC T19381;

DE 04-JUL-1996 (first entry)

DE Human gene signature HUMGS00405

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W09514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATSU) MATSUBARA K.

PA (OKUBA) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1; Page 372; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 119 BP; 37 A; 34 C; 24 G; 24 T;

alignment\_scores:

Quality: 38.00 Length: 11

Ratio: 3.800 Gaps: 0

Percent Similarity: 90.909 Percent Identity: 72.727

alignment\_block:

US-08-653-294-16 x T19381/rev ..

Align seg 1/1 to reverse of: T19381; from: 1 to: 119

2 ArgLeuLeuLeuArgArgGileLeuLeuArgGtyr 12

|||||:|||||:|||||:|||||

110 AGGCTCTTATTAGACGGTTATTCCTGCTACTAC 78

seq\_name: N\_Geneseq\_36:X13394

seq\_documentation\_block:

ID X13394 standard; DNA; 1332 BP.

AC X13394;

DE 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:457.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1997; U08985.

PR 14-NOV-1997; US-066009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI. INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

PS Claim 1; Page 1690-1691; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

SQ Sequence 1332 BP; 479 A; 190 C; 227 G; 431 T;

alignment\_scores:

Quality: 38.00 Length: 12

Ratio: 3.800 Gaps: 0

Percent Similarity: 83.333 Percent Identity: 58.333

alignment\_block:

US-08-653-294-16 x X13394 ..

Align seg 1/1 to: X13394 from: 1 to: 1332

1 TyrArgLeuLeuLeuArgGileLeuLeuArgGtyr 12

|||||:|||||:|||||:|||||

801 TATCGTATTATTATGAGAAAAATACTTATTAAAGGTAT 836

seq\_name: N\_Geneseq\_36:X04340

seq\_documentation\_block:

ID X04340 standard; DNA; 1966 BP.

AC X04340;

DE 13-APR-1999 (first entry)

DE Human secreted protein gene 30 clone HTLAB25.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN W09856804-A1.

PD 17-DEC-1998.

PF 11-JUN-1998; U12125.

PR 02-OCT-1997; US-061060.

PR 13-JUN-1997; US-049547.

PR 13-JUN-1997; US-049548.

PR 13-JUN-1997; US-049549.

PR 13-JUN-1997; US-049550.

PR 13-JUN-1997; US-049606.

PR 13-JUN-1997; US-049607.

PR 13-JUN-1997; US-049608.

PR 13-JUN-1997; US-049609.

PR 13-JUN-1997; US-049610.

PR 13-JUN-1997; US-049611.

PR 13-JUN-1997; US-050566.

PR 13-JUN-1997; US-050901.

PR 13-JUN-1997; US-052989.





AC 051556;  
 DT 18-MAY-1994 (first entry)  
 DE Sequence encoding protein L.  
 KW Protein; immunoglobulin; binding; immobilisation; light chains;  
 KW antibodies; diagnosis; pharmaceutical; ss.  
 OS Peptococcus magnus.  
 FH Key Location/Qualifiers  
 FT 103..3185  
 FT /\*tag= a  
 FT /product= Protein L.  
 FT 490..573  
 FT /\*tag= b  
 FT repeat\_unit  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 673 and 856"  
 FT 574..672  
 FT /\*tag= c  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide position 757"  
 FT 949..1044  
 FT /\*tag= d  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 1162, 1375  
 FT and 1597"  
 FT 1045..1158  
 FT /\*tag= e  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 1261, 1483  
 FT and 1705"  
 FT 1822..1938  
 FT /\*tag= f  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 2347 and 2545"  
 FT 1939..2007  
 FT /\*tag= g  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 2479, 2665  
 FT and 2851"  
 FT 2035..2094  
 FT /\*tag= h  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide position 2209"  
 FT 2095..2208  
 FT /\*tag= i  
 FT /note= "Repeat units are not adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 2269"  
 FT 2914..2934  
 FT /\*tag= j  
 FT /note= "Repeat units are adjacent, repetitions  
 FT of this sequence are not 100% homologous and  
 FT begin at nucleotide positions 2935, 2953,  
 FT 2968, 2986, 3001, 3019 and 3034"  
 PN W09322438-A.  
 PD 11-NOV-1993.  
 PF 07-MAY-1993; G00949.  
 PR (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.  
 PA Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;  
 DR WPI; 93-368797/46.  
 DR Immunoglobulin binding polypeptide, protein L - used for prodn.  
 PT of pharmaceuticals and for immobilising antibodies e.g. on  
 PT columns, in diagnostic tests and in assays  
 PS Disclosure; Figure 1: 29pp; English.  
 CC Protein L forms a complex with immunoglobulin kappa light chain.  
 CC Purified protein can be used as a reagent for immobilising

CC antibodies e.g. on columns, in diagnostic tests and in assays. It  
 CC may also be used in the production of pharmaceuticals.  
 SQ Sequence 3279 BP; 1505 A; 480 C; 626 G; 668 T;  
 alignment\_scores:  
 Quality: 38.00 Length: 12  
 Ratio: 3.167 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 50.000  
 alignment\_block:  
 US-08-653-294-16 x Q51556/rev ..  
 Align seg 1/1 to reverse of: Q51556 from: 1 to: 3279  
 1 TyrArgLeuLeuLeuArgArgLeuLeuLeuArgTyr 12  
 2746 TTTGCGCTGCTTTTTCAGCGTGTGCTTTTAAGATT 2711  
 seq\_name: N\_Geneseq\_36:V74278  
 seq\_documentation\_block:  
 ID V74278 standard; cDNA; 1564 BP.  
 AC V74278;  
 DT 10-MAY-1999 (first entry)  
 DE Arabidopsis Golgi glycosyltransferase clone gtl6.  
 KW Glycosyltransferase; Golgi; gtl6 gene; transgenic plant;  
 KW glycosylation; plant cell wall; paper; pulp; food; ss.  
 OS Arabidopsis thaliana.  
 FH Key Location/Qualifiers  
 FT CDS 100..1449  
 FT /\*tag= a  
 PN W09901558-A1.  
 PD 14-JAN-1999.  
 PF 30-JUN-1998; G01911.  
 PR 30-JUN-1997; GB-013852.  
 PA (UTCA-) UNIV CAMBRIDGE TECH SERVICES LTD.  
 PI Dupree P, Mogelsvang S;  
 DR WPI; 99-106058/09.  
 DR P-PSDB; W90188  
 DT New isolated plant glycosyltransferase nucleic acids - used to  
 PT transform plants to alter the properties or composition of cell  
 PT walls, e.g. for modifying foodstuffs or wood pulp products  
 PS Claim 8; Fig 1A-C; 66pp; English.  
 CC This is the nucleotide sequence of clone gtl6, which encodes a  
 CC novel glycosyltransferase (see W90188) of Arabidopsis thaliana.  
 CC 3 A. thaliana expressed sequence tag (EST) clones were identified  
 CC when searching an EST database with the Schizosaccharomyces pombe  
 CC gma12p and the related Saccharomyces cerevisiae mnl10p proteins.  
 CC An iterative process was then performed to isolate further clones,  
 CC and 5 different genes were identified, i.e. gtl1 (gma twelve-like  
 CC 1), gtl2, gtl4, gtl5 and gtl6 (see V74278-83). The longest clone,  
 CC gtl6, identified from EST H36383, encoded a predicted truncated  
 CC protein, and 5'RACE was performed using RNA from A. thaliana callus  
 CC cells to obtain the full-length clone. The predicted protein has  
 CC the characteristics of a Golgi glycosyltransferase, i.e. a short  
 CC cytosolic N-terminus, single transmembrane domain, and a larger  
 CC luminal C-terminal domain with conserved cysteines. The plant  
 CC glycosyltransferases of the invention (see W90188-94) can catalyse  
 CC one or more steps in the biosynthesis of complex non-cellulosic  
 CC plant cell wall polysaccharides. They may also catalyse the  
 CC glycosylation of plant cell wall protein and/or the transfer of  
 CC saccharide moieties from nucleotide sugar precursors to  
 CC oligosaccharide precursors. Nucleic acids encoding the  
 CC glycosyltransferases can be used to transform plants to alter the  
 CC properties or composition of the cell wall. The plants produced  
 CC can be used in the preparation of e.g. food or wood pulp products  
 CC such as paper or cardboard.  
 SQ Sequence 1564 BP; 474 A; 306 C; 350 G; 434 T;  
 alignment\_scores:  
 Quality: 37.00 Length: 10



```

seq_documentation_block:
ID T79668 standard; cDNA: 7240 BP.
AC T79668:
DE 02-FEB-1998 (first entry)
DI Partial BRCA2 cancer susceptibility gene.
DE BRCA2 cancer susceptibility gene; breast
KW gene therapy; prostate cancer; colorectal cancer; ocular melanoma;
KW leukaemia; human; ss.
OS Homo sapiens.
FH key Location/Qualifiers
FT CDS
FT 2..6989
FT /*tag= a
FT /product= BRCA2 fragment
FT /transl_except= (pos:3800..3801,aa:Glu)
FT /transl_except= (pos:3802..3803,aa:Lys)
PN GB2307477-A.
PD 28-MAY-1997.
PF 25-NOV-1996; 024453.
PR 28-AUG-1996; GB-017961.
PR 23-NOV-1995; GB-023959.
PR 14-DEC-1995; GB-025555.
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PA (YUDU-) UNIV DUKE.
PI Ashworth A, Futreal PA, Stratton MR, Wooster RF;
DR P-PSDB; W25038.
PT Nucleic acid molecules comprising part or all of the BRCA2 cancer
PT susceptibility gene - useful for diagnosis, prognosis or therapeutic
PT treatment of cancer
PS Claim 3; Fig 4; 124pp; English.
CC The present sequence represents the partial sequence of the BRCA2
CC cancer susceptibility gene. The nucleic acid molecule can be used to
CC construct probes for screening cDNA or genomic libraries, sequencing
CC positive clones obtained, and assembling the full length BRCA2
CC sequence. The BRCA2 nucleic acid molecules and proteins are
CC useful in a method of medical treatment, preferably gene therapy,
CC especially for treating cancer, where the cancer is female or male
CC breast cancer, ovarian, prostate or colorectal cancer, ocular melanoma
CC or leukaemia. In particular antisense oligonucleotides capable of
CC hybridising to the BRCA2 nucleic acid, pre-mRNA or mature mRNA are used
CC so that the expression of the BRCA2 nucleic acid is reduced or
CC prevented. The nucleic acid molecules are also useful in a method for
CC diagnosing susceptibility or predisposition to cancer in a patient. The
CC nucleic acid molecules are used to design probes or primers for PCR to
CC determine or detect the presence of mutations in a sample of nucleic
CC acid from a patient. The BRCA2 promoter region is useful for screening
CC for substances which modulate the expression of nucleic acid under
CC control of the promoter. Antibodies are used to determine the presence,
CC amount or location in a cell of a BRCA2 polypeptide or its mutant forms.
CC The polypeptides are used to screen for binding partners, these are
CC useful to screen for substances which mimic the activity of BRCA2
CC polypeptide, which can be used as cancer therapeutics.
SQ Sequence 7240 BP; 2735 A; 1188 C; 1322 G; 1995 T;

alignment_scores:
Quality: 37.00 Length: 11
Ratio: 3.700 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 54.545

alignment_block:
US-08-653-294-16 x T79668 ..
Align seg 1/1 to: T79668 from: 1 to: 7240
2 ArgLeuLeuLeuArgIleLeuLeuArgTyr 12
||||:||||:||||:||||:||||:
2892 AGAATATTAGTAGGAACTTCTGCAGAGGTAC 2924

seq_name: N_Geneseq_36:V52245
seq_documentation_block:
ID V52245 standard; DNA: 7902 BP.
AC V52245:
DE 23-OCT-1998 (first entry)
DI Streptococcus pneumoniae genome fragment SEQ ID NO:112.
DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588
PR 31-OCT-1996; US-029960.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 814-819; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments can be used in S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 7902 BP; 2482 A; 1307 C; 1634 G; 2479 T;

alignment_scores:
Quality: 37.00 Length: 10
Ratio: 3.700 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 60.000

alignment_block:
US-08-653-294-16 x V52245 ..
Align seg 1/1 to: V52245 from: 1 to: 7902
1 TyrArgLeuLeuLeuArgIleLeuLeu 10
||||:||||:||||:||||:||||:
7791 TTTCGTATGCTAGTTCGAGATTGTGCTA 7820

seq_name: N_Geneseq_36:Q79140
seq_documentation_block:
ID Q79140 standard; cDNA: 9487 BP.
AC Q79140:
DE 07-SEP-1995 (first entry)
DI Hepatitis C virus gene HC-G9 cDNA.
DE Hepatitis C virus; HCV; non-A non-B; HC-G9; treatment; ss.
KW Hepatitis C virus.
OS Hepatitis C virus.
FH key Location/Qualifiers
FT cds
FT 342..9377
FT /*tag= a
FT /transl_except= pos:7947..7949, aa:Glu
FT /transl_except= pos:7950..7952, aa:Glu
FT /transl_except= pos:7953..7955, aa:Ala

```

```
FT /transl_except- pos:7956..7958, aa:Cys
FT /transl_except- pos:7959..7961, aa:Ser
FT /transl_except- pos:7962..7964, aa:Leu
FT /transl_except- pos:7965..7967, aa:Thr
FT /transl_except- pos:7968..7970, aa:Pro
FT /transl_except- pos:7971..7973, aa:Pro
FT /transl_except- pos:7974..7976, aa:His
FT /transl_except- pos:7977..7979, aa:Ser
FT /transl_except- pos:7980..7982, aa:Ala
FT /transl_except- pos:7983..7985, aa:Arg
FT /transl_except- pos:7986..7988, aa:Ser
FT /transl_except- pos:7989..7991, aa:Lys
FT /transl_except- pos:7992..7994, aa:Tyr
FT /transl_except- pos:7995..7997, aa:Asp
FT /transl_except- pos:7998..8000, aa:Val
FT /transl_except- pos:8001..8003, aa:Val
FT /transl_except- pos:8004..8006, aa:Lys
FT /transl_except- pos:8007..8009, aa:Lys
FT /transl_except- pos:8010..8012, aa:Leu
FT /transl_except- pos:8013..8015, aa:Pro
FT /transl_except- pos:8016..8018, aa:Ile
FT /transl_except- pos:8019..8021, aa:Ala
FT /transl_except- pos:8022..8024, aa:Val
FT /transl_except- pos:8025..8027, aa:Met
FT /transl_except- pos:8028..8030, aa:Gly
FT /transl_except- pos:8031..8033, aa:Thr
FT /transl_except- pos:8034..8036, aa:Ser
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO ) IMMUNO JAPAN KK.
DR WPI: 95-040318/06.
DR P-PSDB: R67588.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
PT treatment of hepatitis C
PS Claim 1; Pages 10-15; 41pp; Japanese.
CC Q79140 is the hepatitis C virus (HCV) gene HC-C9 cDNA, it encodes
CC the protein described in R67588. Both the cDNA and protein can be
CC used in the treatment of HCV infection
SQ Sequence 9487 BP; 1905 A; 2798 C; 2687 G; 2097 T;
```

```
alignment_scores:
  Quality: 37.00 Length: 12
  Ratio: 3.700 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 58.333
```

alignment\_block:

US-08-653-294-16 x Q79140/rev ..

Align seg 1/1 to reverse of: Q79140 from: 1 to: 9487

1 TyrArgLeuIleArgArgIleLeuLeuArgTyr 12

|||||:|||||:|||||:|||||:|||||

2597 TATCAAGTTCACAGCGCGCTCTACCTGAGATAT 2562

seq\_name: N\_Geneseq\_36:X30256

seq\_documentation\_block:

ID X30256 standard; cDNA; 10485 BP.

AC X30256;

DT 21-JUN-1999 (first entry)

DE Human BRCA2 (om12) encoding cDNA.

KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;

KW gene therapy; breast cancer; ovarian cancer; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 229..10485

FT /\*tag= a

PN W09909164-A1.

PD 25-FEB-1999.

PF 14-AUG-1998; U16905.

```
PR 22-MAY-1998; US-084471.
PR 15-AUG-1997; US-055784.
PR 07-NOV-1997; US-064926.
PR 12-NOV-1997; US-065367.
PR 01-MAY-1998; US-071715.
PA (ONCO-) ONCORMED INC.
PI Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M,
PI Rabin MB, Schryer B, White MB, Yoshikawa M;
DR WPI: 99-190163/16.
DR P-PSDB: Y04355.
PT New coding sequence haplotypes of the human BRCA2 gene - used to
PT develop products for determining susceptibility to, detection and
PT treatment of breast or ovarian cancer
PS Claim 12; Page 75-90; 226pp; English.
CC The present invention describes genomic DNA which contains a BRCA2 gene
CC where the first 12 nucleotides beginning exon 5 are 5'-TCCCTGTTGTTCT-3'
CC as in sequence (I) (see X03249), where nucleotide numbers
CC 5782-5790 are GTTGTGTT as in sequence (IV) (see X03255), and where the
CC last 20 nucleotides encoding exon 15 are 5'-CTGCGGTGTTCTATAACAG-3' as
CC in sequence (II) (see X03251) and the first 20 nucleotides beginning
CC exon 16 are 5'-CTGTATAGTATGGGTTTC-3' as in sequence (III) (see X03253).
CC products and methods from the present invention can be used for
CC identifying mutations in the BRCA2 gene leading to predisposition or
CC higher susceptibility to breast or ovarian cancer. They can also be used
CC for detection and gene therapy for breast and ovarian cancers. They can
CC be used in methods for monitoring disease progression, for determining
CC patients suited for gene and protein replacement progression, or for
CC detecting the presence or quantifying the amount of a tumour growth
CC inhibitor following such therapy. The BRCA2 protein, polypeptides, their
CC functional equivalents, antibodies, and PMS may also be useful in the
CC study of the characteristics of BRCA2 proteins, such as structure and
CC function of BRCA2 in oncogenesis or subcellular localisation of BRCA2
CC protein in normal and cancerous cells. The present sequence encodes
CC human BRCA2 (om12).
SQ Sequence 10485 BP; 3810 A; 1842 C; 1962 G; 2870 T;
```

alignment\_scores:

Quality: 37.00 Length: 11

Ratio: 3.700 Gaps: 0

Percent Similarity: 90.909 Percent Identity: 54.545

alignment\_block:

US-08-653-294-16 x X30256 ..

Align seg 1/1 to: X30256 from: 1 to: 10485

2 ArgLeuLeuIleArgArgIleLeuLeuArgTyr 12

|||||:|||||:|||||:|||||

3962 AGAATATTAGTAGGAAACTTCTGCAGAGGTAC 3994

seq\_name: N\_Geneseq\_36:X30257

seq\_documentation\_block:

ID X30257 standard; cDNA; 10485 BP.

AC X30257;

DT 21-JUN-1999 (first entry)

DE Human BRCA2 (om13) encoding cDNA.

KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;

KW gene therapy; breast cancer; ovarian cancer; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 229..10485

FT /\*tag= a

PN W09909164-A1.

PD 25-FEB-1999.

PF 14-AUG-1998; U16905.

PR 22-MAY-1998; US-084471.

PR 15-AUG-1997; US-055784.

PR 07-NOV-1997; US-064926.

PR 12-NOV-1997; US-065367.

PR 01-MAY-1998; US-071715.

PA (ONCO-) ONCORMED INC.

```

PI Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M,
PI Rabin MB, Schryer B, White MB, Yoshikawa M;
DR WPI: 99-190163/16.
DR P-PSDB: Y04356.
PT New coding sequence haplotypes of the human BRCA2 gene - used to
PT develop products for determining susceptibility to, detection and
PT treatment of breast or ovarian cancer
PS Claim 13: Page 97-111; 226pp; English.
CC The present invention describes genomic DNA which contains a BRCA2 gene
CC as in sequence (I) (see X30249), where nucleotides numbers
CC 5782-5790 are GTTGTGTT as in sequence (IV) (see X30255), and where the
CC last 20 nucleotides encoding exon 15 are 5'-CTGGGTGTCTCAAAACAG-3' as
CC in sequence (II) (see X30251) and the first 20 nucleotides beginning
CC exon 16 are 5'-CTGTATACGTATGCGTTC-3' as in sequence (III) (see X30253).
CC Products and methods from the present invention can be used for
CC identifying mutations in the BRCA2 gene leading to predisposition or
CC higher susceptibility to breast or ovarian cancer. They can also be used
CC for detection and gene therapy for breast and ovarian cancers. They can
CC be used in methods for monitoring disease progression, for determining
CC patients suited for gene and protein replacement progression, or for
CC detecting the presence or quantifying the amount of a tumour growth
CC inhibitor following such therapy. The BRCA2 protein, polypeptides, their
CC functional equivalents, antibodies, and PNs may also be useful in the
CC study of the characteristics of BRCA2 proteins, such as structure and
CC function of BRCA2 in oncogenesis or subcellular localisation of BRCA2
CC protein in normal and cancerous cells. The present sequence encodes
CC human BRCA2 (om13).
SQ Sequence 10485 BP; 3808 A; 1842 C; 1965 G; 2870 T;

alignment_scores:
Quality: 37.00 Length: 11
Ratio: 3.700 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 54.545

alignment_block:
US-08-653-294-16 x X30257 ..

Align seg 1/1 to: X30257 from: 1 to: 10485

2 ArgLeuLeuIleArgIleLeuLeuArgTyr 12
||||:||||:||||:||||:||||:
3962 AGAATATTAGTACGAGAACTTCTGCAGAGGTAC 3994

seq_name: N_Geneseq_36:X30258

seq_documentation_block:
ID X30258 standard; cDNA; 10485 BP.
AC X30258;
DE 21-JUN-1999 (first entry)
DE Human BRCA2 (om14) encoding cDNA.
KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;
KW gene therapy; breast cancer; ovarian cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 229..10485
FT /*tag= a
FT WO9909164-A1.
PN 25-FEB-1999.
PD 14-AUG-1998; U16905.
PR 22-MAY-1998; US-084471.
PR 15-AUG-1997; US-055784.
PR 07-NOV-1997; US-064926.
PR 12-NOV-1997; US-065367.
PR 01-MAY-1998; US-071715.
PA (ONCO-) ONCORMED INC.
PI Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M,
PI Rabin MB, Schryer B, White MB, Yoshikawa M;
DR WPI: 99-190163/16.
DR P-PSDB: Y04357.
PT New coding sequence haplotypes of the human BRCA2 gene - used to
PT develop products for determining susceptibility to, detection and

```

```

PT treatment of breast or ovarian cancer
PS Claim 14: Page 119-133; 226pp; English.
CC The present invention describes genomic DNA which contains a BRCA2 gene
CC where the first 12 nucleotides beginning exon 5 are 5'-TCCGTGTGTTCT-3'
CC as in sequence (I) (see X03249), where nucleotides numbers
CC 5782-5790 are GTTGTGTT as in sequence (IV) (see X30255), and where the
CC last 20 nucleotides encoding exon 15 are 5'-CTGGGTGTCTCAAAACAG-3' as
CC in sequence (II) (see X30251) and the first 20 nucleotides beginning
CC exon 16 are 5'-CTGTATACGTATGCGTTC-3' as in sequence (III) (see X30253).
CC Products and methods from the present invention can be used for
CC identifying mutations in the BRCA2 gene leading to predisposition or
CC higher susceptibility to breast or ovarian cancer. They can also be used
CC for detection and gene therapy for breast and ovarian cancers. They can
CC be used in methods for monitoring disease progression, for determining
CC patients suited for gene and protein replacement progression, or for
CC detecting the presence or quantifying the amount of a tumour growth
CC inhibitor following such therapy. The BRCA2 protein, polypeptides, their
CC functional equivalents, antibodies, and PNs may also be useful in the
CC study of the characteristics of BRCA2 proteins, such as structure and
CC function of BRCA2 in oncogenesis or subcellular localisation of BRCA2
CC protein in normal and cancerous cells. The present sequence encodes
CC human BRCA2 (om14).
SQ Sequence 10485 BP; 3809 A; 1841 C; 1964 G; 2871 T;

alignment_scores:
Quality: 37.00 Length: 11
Ratio: 3.700 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 54.545

alignment_block:
US-08-653-294-16 x X30258 ..

Align seg 1/1 to: X30258 from: 1 to: 10485

2 ArgLeuLeuIleArgIleLeuLeuArgTyr 12
||||:||||:||||:||||:||||:
3962 AGAATATTAGTACGAGAACTTCTGCAGAGGTAC 3994

```

OM of: US-08-653-294-16 to: EST.\* out\_format : pfs

Date: Feb 8, 2000 4:03 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp  
-O=Cgml\_1/USPTO\_spool/US08653294/runat\_04022000\_160700\_15770/app\_query.fasta.1  
-DB=EST -QFMT=fastac -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000  
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-16

Query length: 12

Database: EST.\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 8553.360000

score\_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_est36:AV182651	+	45.00	149.71	14.22	300	AV182651 AV182651 Yuji Kohara
gb_gss9:AQ137850	+	43.00	139.21	54.70	461	AQ137850 HS_3058_B2_C09_MF CIT
gb_gss6:AQ835306	+	43.00	136.95	73.10	605	AQ835306 HS_4815_A1_D12_T7A CIT
gb_est19:AA748739	-	42.00	139.07	55.68	309	AA748739 ny06e12.s1 NCI-CGAP-GC
gb_gss10:AQ253565	-	42.00	135.24	91.05	439	AQ253565 HS_3026_A1_A07_MF CIT
gb_gss19:AA103036	-	42.00	134.76	96.80	519	AA103036 mo21d11.r1 Life Tech
gb_gss14:AQ569045	-	42.00	133.61	112.19	596	AQ569045 HS_5344_A1_F08_T7A RPO
gb_gss11:AQ259553	+	42.00	132.45	130.14	695	AQ259553 nbxb0023E12r CUGI Rice
gb_est11:D37193	+	41.00	134.34	102.21	360	D37193 CELK04106F Yuji Kohara
gb_gss11:AQ111590	+	41.00	133.06	120.47	420	AQ111590 RPI111-1A13.r1 RPI-CIT-11
gb_gss6:AQ869567	-	41.00	132.71	125.98	438	AQ869567 nbcb0035D05f CUGI Rice
gb_est37:AI960153	-	41.00	132.35	131.82	457	AI960153 sc39b09.x1 Gm-cl014 G1
gb_est33:AI774609	+	41.00	130.40	169.34	593	AI774609 EST255709 tomato resis
gb_gss3:B70241	+	41.00	130.19	174.03	578	B70241 CIT-HSP-2028I12 TR CIT-H
gb_est33:AI775248	+	41.00	129.91	180.29	613	AI775248 EST256348 tomato resis
gb_est38:AW096409	+	41.00	129.61	187.51	636	AW096409 EST289589 tomato mixed
gb_est24:AI162936	+	40.00	133.84	108.97	252	AI162936 A07P320 Hybrid aspen
gb_est41:AV263447	+	40.00	133.42	114.97	265	AV263447 AV263447 RIKEN full-le
gb_gss9:AQ097888	+	40.00	131.65	144.34	328	AQ097888 HS_3036_B2_A06_MF CIT
gb_gss3:B60584	-	40.00	131.08	155.15	351	B60584 CIT-HSP-2013F20 TPC CIT-
gb_est37:AI960040	-	40.00	129.85	181.68	407	AI960040 sc37c08.x1 Gm-cl014 G1
gb_est24:AI163552	+	40.00	129.75	184.06	422	AI163552 A04P300 Hybrid aspen
gb_est39:AQ099131	+	40.00	129.26	196.00	437	AQ099131 SG34H05.y1 Gm-cl012 G1
gb_est24:AI163545	+	40.00	129.20	197.43	440	AI163545 A04P230 Hybrid aspen
gb_est37:AI966327	+	40.00	128.96	203.66	453	AI966327 sc37c08.y1 Gm-cl014 G1
gb_est18:AA690244	-	40.00	128.85	206.54	459	AA690244 vt82g11.s1 Soares mous
gb_gss13:AQ476031	+	40.00	128.55	214.70	476	AQ476031 CITBI-EL-2591A11 TF CIT
gb_est8:AAQ12276	-	40.00	128.36	220.00	487	AAQ12276 T8EST2216c11.r1 T9ME49
gb_gss12:AQ339784	+	40.00	128.23	223.85	495	AQ339784 HS_5021_B1_E02_T7 RPI-CIT
gb_est24:AI161741	+	40.00	127.80	236.41	521	AI161741 A06P200 Hybrid aspen
gb_gss8:AQ055939	-	40.00	127.77	237.38	523	AQ055939 CIT-HSP-2340F9 TF CIT-
gb_est24:AI162058	+	40.00	127.75	237.86	524	AI162058 A011P530 Hybrid aspen
gb_est24:AI161931	+	40.00	127.64	241.25	531	AI161931 A009P660 Hybrid aspen
gb_est24:AI162067	+	40.00	127.50	245.62	540	AI162067 A011P630 Hybrid aspen
gb_est37:AI966443	+	40.00	126.81	268.48	587	AI966443 sc39b08.y1 Gm-cl014 G1
gb_est24:AI161739	+	40.00	126.78	269.45	589	AI161739 A06P180 Hybrid aspen
gb_est44:AW219466	+	40.00	126.50	279.22	609	AW219466 EST301864 tomato root
gb_gss11:AQ328194	-	40.00	126.22	289.50	630	AQ328194 nbxb0042L03r CUGI Rice
gb_gss6:AQ840209	-	40.00	125.75	307.66	667	AQ840209 nbxb0052015f CUGI Rice
gb_gss11:AQ291628	-	40.00	123.80	394.93	843	AQ291628 nbxb0039C22r CUGI Rice
gb_est18:AA709270	-	39.00	131.66	144.19	216	AA709270 a121e10.s1 Soares-test
gb_est22:AU006893	-	39.00	128.62	212.68	311	AU006893 AU006893 Schizosacchar

gb\_est36:AI927435 + 39.00 128.49 216.33 316 ! AI927435 w075f12.x1 NCI-CGAP  
gb\_est16:AA591962 + 39.00 127.96 231.69 337 ! AA591962 v150b07.r1 Beddingt  
gb\_gss15:AQ632742 + 39.00 127.34 250.80 363 ! AQ632742 RPCI-11-475D13.TJ R

seq\_name: gb\_est36:AV182651

seq\_documentation\_block:

LOCUS AV182651 300 bp mRNA EST 21-JUL-1999

DEFINITION AV182651 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite  
embryo Caenorhabditis elegans cDNA clone yk640g7 3', mRNA sequence.

ACCESSION AV182651

VERSION AV182651.1 GI:5562552

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 300)

AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,  
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,  
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and  
Nomoto,H.

TITLE Expressed genes in C.elegans

JOURNAL Unpublished (1999)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3189500.

CONTACT: Yuji Kohara

Gene Library Lab

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

Location/Qualifiers

1..300

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk640g7"

/clone.lib="Yuji Kohara unpublished cDNA:Strain N2

hermaphrodite embryo"

/sex="hermaphrodite"

/dev\_stage="embryo"

54 g

68 c

71 t

2 others

BASE COUNT 105 a

ORIGIN

alignment\_scores:

Quality: 45.00 Length: 12

Ratio: 4.091 Gaps: 0

Percent Similarity: 91.667 Percent Identity: 66.667

alignment\_block:

US-08-653-294-16 x AV182651 ..

Align seg 1/1 to: AV182651 from: 1 to: 300

1 TyrArgLeuLeuArgArgArgLeuLeuLeuArgTyr 12

|||||

85 TACCGACTACTTCGCTAGTATTATTAACATAT 120

seq\_name: gb\_gss9:AQ137850

seq\_documentation\_block:

LOCUS AQ137850 461 bp DNA GSS 24-SEP-1998

DEFINITION HS\_3038\_B2\_C09\_MF CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate-3058 Col-18 Row-F, genomic survey

sequence.

ACCESSION AQ137850

VERSION AQ137850.1 GI:3528503

KEYWORDS GSS.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 461)  
**REFERENCE**  
**AUTHORS** Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
**JOURNAL** Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
**MEDLINE** 99380589  
**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
 High throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 3058 row: F column: 18  
 Class: BAC ends  
 High quality sequence stop: 461

**FEATURES**  
 source  
 1..461  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-3058 Col-18 Row-F"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

**BASE COUNT** 132 a 125 c 102 g 102 t  
**ORIGIN**

**alignment\_scores:**  
 Quality: 43.00 Length: 12  
 Ratio: 4.300 Gaps: 0  
 Percent Similarity: 83.333 Percent Identity: 75.000

**alignment\_block:**  
 US-08-653-294-16 x A0137850 ..  
 Align seg 1/1 to: A0137850 from: 1 to: 461

1 TyrArgLeuLeuIleArgArgIleLeuLeuArgTyr 12  
 :::::::::::::::::::::::::::::::  
 18 CATCGACTCACATATAGGCGAATCTGCCAGGTAC 53

**seq\_name:** gb\_gss6:A0835306

**seq\_documentation\_block:**  
**LOCUS** A0835306 605 bp DNA GSS 27-AUG-1999  
**DEFINITION** HS\_4815\_AL\_D12\_T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-4815 Col-23 Row-G, genomic survey sequence.  
**ACCESSION** A0835306  
**VERSION** A0835306.1 GI:5801368  
**KEYWORDS** GSS.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
**AUTHORS** Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
**JOURNAL** Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
**MEDLINE** 99380589  
**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 4815 row: G column: 23  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 605.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..605  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-4815 Col-23 Row-G"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

**BASE COUNT** 178 a 112 c 107 g 201 t 7 others  
**ORIGIN**

**alignment\_scores:**  
 Quality: 43.00 Length: 12  
 Ratio: 3.909 Gaps: 0  
 Percent Similarity: 91.667 Percent Identity: 58.333

**alignment\_block:**  
 US-08-653-294-16 x A0835306 ..  
 Align seg 1/1 to: A0835306 from: 1 to: 605

1 TyrArgLeuLeuIleArgArgIleLeuLeuArgTyr 12  
 :::::::::::::::::::::::::::::::  
 477 TATACGCTCTTATGAGGAGGCTCTCATCAATAT 512

**seq\_name:** gb\_est19:AA748739

**seq\_documentation\_block:**  
**LOCUS** AA748739 309 bp mRNA EST 22-JAN-1998  
**DEFINITION** ny06e12.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1270990 3', mRNA sequence.  
**ACCESSION** AA748739  
**VERSION** AA748739.1 GI:2788697  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
**AUTHORS** NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** On Jan 14, 1998 this sequence version replaced gi:1797612.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Cloned distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 674 Std Error: 0.00  
 Seq primer: -40m13 fwd. Bt from Amersham  
 High quality sequence stop: 256.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..309



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_GCB1"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+ IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      95 a      55 g      107 t
ORIGIN

alignment_scores:
  Quality: 42.00      Length: 12
  Ratio: 4.200      Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667

alignment_block:
US-08-653-294-16 x AA748739/rev ..
Align seg 1/1 to reverse of: AA748739 from: 1 to: 309

1 TyrArgLeuLeuLeuArgArgileLeuLeuArgTyr 12
||||| :|||||:|||||:|||||:|||||
278 TACCGAGATGATACGACGTGACTGCTCTCTAC 243

seq_name: gb_gss10:AQ253565

seq_documentation_block:
LOCUS AQ253565 490 bp DNA GSS 10-OCT-1998
DEFINITION HS_3026_A1_A07_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3026 Col=13 Row=A, genomic survey
sequence.
ACCESSION AQ253565
VERSION AQ253565.1 GI:3724919
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 490)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
PROC. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 9980589
CONTACT: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Sequence Tagged Connector
Plate: 3026 row: A column: 13
Class: BAC ends
High quality sequence stop: 490.
Location/Qualifiers
1..490
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone="Plate=3026 Col=13 Row=A"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      177 a      99 c      81 g      132 t      1 others
ORIGIN

alignment_scores:
  Quality: 42.00      Length: 10
  Ratio: 4.200      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:
US-08-653-294-16 x AQ253565 ..
Align seg 1/1 to: AQ253565 from: 1 to: 490

1 TyrArgLeuLeuLeuArgArgileLeuLeu 10
|||||:|||||:|||||:|||||:|||||
25 TATCGACTACTACTAGCGGACTTCTATTA 54

seq_name: gb_est9:AA103036

seq_documentation_block:
LOCUS AA103036 519 bp mRNA EST 29-OCT-1996
DEFINITION mo2id11.r1 Life Tech mouse embryo 13 5dpc 10666014 Mus musculus
cDNA clone IMAGE:554229 5', mRNA sequence.
ACCESSION AA103036
VERSION AA103036.1 GI:1649195
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 519)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On May 8, 1995 this sequence version replaced gi:801493.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:335021
Seq primer: -28M13 rev1 from Amersham
High quality sequence stop: 444.
Location/Qualifiers
1..519
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:554229"
/clone_lib="Life Tech mouse embryo 13 5dpc 10666014"
/tissue_type="embryo"
/dev_stage="13.5dpc embryos"
/lab_host="DH10B"
/note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. 13.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT      150 a      131 c      128 g      110 t
ORIGIN

```



Percent Similarity: 83.333 Percent Identity: 75.000

## alignment\_block:

US-08-653-294-16 x AQ259553 ..

Align seg 1/1 to: AQ259553 from: 1 to: 685

1 TyrArgLeuLeuIleArgArgIleLeuLeuArgTyr 12  
|||||.....  
453 TACAGGCTGTAGTCTTATTCATACGTCTACGCTAT 488

seq\_name: gb\_est1:D37193

seq\_documentation\_block:

LOCUS D37193 360 bp mRNA EST 24-OCT-1994  
DEFINITION CELK041G6F Yuji Kohara unpublished cDNA Caenorhabditis elegans cdNA  
clone yk41g6 5', mRNA sequence.

ACCESSION D37193.1 GI:525575

VERSION D37193.1

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and

Tabara, H.

Toward an expression map of the C.elegans genome

Unpublished (1994)

CONTACT Yuji Kohara

Gene Library Lab

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

FEATURES

Location/Qualifiers

1..360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/gb\_xref="taxon:6239"

/clone="yk41g6"

/clone\_lib="Yuji Kohara unpublished cdNA"

/sex="hermaphrodite, male"

/tissue\_type="whole animal"

/dev\_stage="varied"

BASE COUNT 106 a 76 c 80 g 97 t 1 others

ORIGIN

## alignment\_scores:

Quality: 41.00 Length: 12

Ratio: 3.727 Gaps: 0

Percent Similarity: 91.667 Percent Identity: 58.333

## alignment\_block:

US-08-653-294-16 x D37193 ..

Align seg 1/1 to: D37193 from: 1 to: 360

1 TyrArgLeuLeuIleArgArgIleLeuLeuArgTyr 12  
|||||.....  
78 TTCAGACTGCTTGAACAGTGTCTCTTTATTAC 113

seq\_name: gb\_gss11:AQ311590

seq\_documentation\_block:

LOCUS AQ311590 420 bp DNA GSS 04-MAY-1999  
DEFINITION RPC111-1A13.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-1A13,  
genomic survey sequence.

ACCESSION AQ311590

VERSION AQ311590.1

KEYWORDS GI:4043339

GSS.

SOURCE  
ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 420)

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,

Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

Use of human BAC End Sequences for Sequence-Ready Map Building

Unpublished (1998)

Other GSSs: RPC111-1A13.TVB

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeatig@ig.org

Clones are derived from the human BAC library RPC1-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..420

/organism="Homo sapiens"

/db\_xref="GBB:7500012"

/db\_xref="taxon:9606"

/clone="RPC1-11-1A13"

/clone\_lib="RPC1-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

BASE COUNT 133 a 80 c 82 g 125 t

ORIGIN

alignment\_scores:

Quality: 41.00 Length: 10

Ratio: 4.100 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 80.000

alignment\_block:

US-08-653-294-16 x AQ311590 ..

Align seg 1/1 to: AQ311590 from: 1 to: 420

3 LeuLeuIleArgArgIleLeuLeuArgTyr 12

|||||.....

382 CTCTCTTTAGAGGGCTCTTCAGATAT 411

seq\_name: gb\_gss6:AQ869567

seq\_documentation\_block:

LOCUS AQ869567 438 bp DNA GSS 03-NOV-1999

DEFINITION nbe0035D03f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic

clone nbe0035D03f, genomic survey sequence.

ACCESSION AQ869567

VERSION AQ869567.1

KEYWORDS GI:6220018

GSS.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Oryza.

1 (bases 1 to 438)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

JOURNAL

COMMENT On Mar 23, 1999 this sequence version replaced gi:3324665.

Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATAGCACTCACTATAGGG  
Class: BAC ends  
High quality sequence start: 24  
High quality sequence stop: 397.

#### FEATURES

Location/Qualifiers  
1..438  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbe0035D03f"  
/tissue\_type="Leaf"  
/lab\_host="E. coli DH10B"  
/note="Vector: pBACIndigo; Site\_1: EcoRI; Site\_2: EcoRI;  
Rice is the most important food crop in the world. Half of  
the world population, especially those inhabiting highly  
populated areas of the humid tropics and subtropics, rely  
on rice as their primary source of carbohydrate.  
Monocotyledonous rice is a diploid plant (2n=24) with a  
haploid genome equivalent of 431 Mbp (Arumuganathan and  
Earle, 1991). The relatively small genome of rice, three  
times larger than that of Arabidopsis, makes it suitable  
for genomic studies. In order to facilitate positional  
cloning, physical mapping and genome sequencing of rice,  
we have constructed a BAC library from Oryza sativa.  
Nipponbare variety using EcoRI as the cloning enzyme. The  
library contains 55,296 clones with an average insert size  
of 121 kb providing approximately 15 haploid genome  
equivalents. The deep coverage allows the isolation a  
particular sequence with a probability of 99.9 %. Three  
high density filters, each containing 18,432 clones  
(doubly spotted), represent the whole library for colony  
screening and can be requested from the Clemson University  
BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 121 a 86 c 65 g 164 t 2 others  
ORIGIN  
alignment\_scores:  
Quality: 41.00 Length: 10  
Ratio: 4.536 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 90.000

#### alignment\_block:

US-08-653-294-16 x AQ869567 ..  
Align seg 1/1 to: AQ869567 from: 1 to: 438

3 LeuLeuLeuArgArgIleLeuLeuArgTyr 12  
|||||  
15 CTCACTATACGGCAATTCATTAAAGATAT 44

seq\_name: gb\_est37:AI960153

seq\_documentation\_block:  
LOCUS AI960153 457 bp mRNA EST 20-AUG-1999  
DEFINITION sc39b09.x1 Gm-cl014 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl014-1458 3', mRNA sequence.  
ACCESSION AI960153  
VERSION AI960153.1 GI:5752866  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core

#### REFERENCE

##### AUTHORS

1. (bases 1 to 457)  
Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,  
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,  
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,  
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,  
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,  
McCann,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)

#### TITLE

##### JOURNAL

##### COMMENT

On Mar 10, 1998 this sequence version replaced gi:2948010.  
Other ESTs: sc39b09.y1  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Seq primer: primer name ambiguous  
High quality sequence stop: 364.

#### FEATURES

##### source

Location/Qualifiers  
1..457  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl014-1458"  
/clone\_lib="Gm-cl014"  
/tissue\_type="Leaves, 2-3 week old seedlings, greenhouse  
grown"

/lab\_host="DH10B"  
/note="Vector: p7T3Pac (p7T3, Pharmacia); Site\_1: EcoRI;  
Site\_2: HindIII; This cDNA library was constructed from  
mRNA isolated from leaves of 2-3 week old greenhouse grown  
plants. Complementary DNA was synthesized from mRNA using  
a 3' anchored poly (dT) primer. EcoRI adapters were  
ligated to the blunt-ended cDNA fragments followed by  
digestion with EcoRI and HindIII. The cDNA fragments were  
directionally cloned into the EcoRI-HindIII restriction  
site of the p7T3-Pac vector. The ligated cDNA fragments  
were transformed into DH10B host cells (Gibco BRL). This  
library was constructed by Dr. Randy Shoemaker and Dr.  
John Erpelding."

BASE COUNT 106 a 106 c 160 g 85 t  
ORIGIN

#### alignment\_scores:

Quality: 41.00 Length: 10  
Ratio: 4.100 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 80.000

#### alignment\_block:

US-08-653-294-16 x AI960153/rev ..

Align seg 1/1 to reverse of: AI960153 from: 1 to: 457

2 ArgLeuLeuArgArgIleLeuLeuArg 11  
|||||  
236 CGCCTCCTCGTCGACGACTTCCTCCGA 207

seq\_name: gb\_est33:AI774609

#### seq\_documentation\_block:

LOCUS AI774609 578 bp mRNA EST 29-JUN-1999  
DEFINITION EST255709 tomato resistant, Cornell Lycopersicon esculentum cDNA  
clone CLER12H12, mRNA sequence.  
ACCESSION AI774609

```

VERSION      AI774609.1  GI:5272650
KEYWORDS     EST.
SOURCE       Lycopersicon esculentum
ORGANISM     tomato.

REFERENCE    1 (bases 1 to 578)
AUTHORS      D' Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
              Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y.,
              Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
              Giovannoni, J.J. and Martin, G.B.
              Generation of ESTs from Pseudomonas resistant tomato
              Unpublished (1999)
              On Dec 20, 1995 this sequence version replaced gi:1136039.
              Other_ESTs: IC2277

TITLE        Contact: David Frisch
JOURNAL      Clemson University Genomics Institute
COMMENT      100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 4366
              Fax: 864 656 4293
              Email: dfrisch@CLEMSON.EDU
              5 prime sequence.

FEATURES     source
              Location/Qualifiers
              1..578
               /organism="Lycopersicon esculentum"
               /cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
               /db_xref="taxon:4081"
               /clone_lib="tomato resistant, Cornell"
               /tissue_type="leaf"
               /dev_stage="4-week old"
               /lab_host="SOLR"
               /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
               XhoI; CLER - Tomato Pseudomonas Resistant EST Library.
               Directionally cloned cDNAs inserted into pBluescript
               SK(-) at 5' end with EcoRI and 3' end with XhoI site."

BASE COUNT   169 a 139 c 133 g 137 t
ORIGIN

alignment_scores:
  Quality: 41.00      Length: 11
  Ratio: 4.556       Gaps: 0
  Percent Similarity: 81.818      Percent Identity: 81.818

alignment_block:
  US-08-653-294-16 x AI774609      ..
  Align seg 1/1 to: AI774609 from: 1 to: 578

1 TyrArgLeuLeuLeuArgArgLeuLeuArg 11
|||||
3 TATCGTCTCTGTGCGCGGTATCTCTACGC 35

seq_name: gb_gss3:B70241

seq_documentation_block:
LOCUS      B70241      593 bp      DNA
DEFINITION CIT-HSP-2028112.TR CIT-HSP Homo sapiens genomic clone 2028112,
genomic survey sequence.
ACCESSION  B70241
VERSION    B70241.1  GI:2709465
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 593)
            Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K.,
            Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,


```

```

TITLE        Simon, M. and Venter, J.C.
JOURNAL      Building
COMMENT      Use of a random BAC End Sequence Database for Sequence-Ready Map
              Unpublished (1997)
              Other_GSSs: CIT-HSP-2028112.TFB
              Contact: Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: mdadams@tigr.org
              Clones are available from Research Genetics (info@resgen.com). BAC
              end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
              Seq primer: M13 Reverse
              Class: BAC ends.

FEATURES     source
              Location/Qualifiers
              1..593
               /organism="Homo sapiens"
               /db_xref="GDB:7048142"
               /db_xref="taxon:9606"
               /clone="2028112"
               /clone_lib="CIT-HSP"
               /sex="Male"
               /cell_type="Sperm"
               /note="vector: pBelOBAC11; Site_1: HindIII; Site_2:
               HindIII"

BASE COUNT   238 a 121 c 127 g 106 t
ORIGIN

alignment_scores:
  Quality: 41.00      Length: 11
  Ratio: 4.100       Gaps: 0
  Percent Similarity: 90.909      Percent Identity: 72.727

alignment_block:
  US-08-653-294-16 x B70241      ..
  Align seg 1/1 to: B70241 from: 1 to: 593

2 ArgLeuLeuLeuArgArgLeuLeuArgTyr 12
|||||
304 AGACATCTTCTTAAGAAGACTACTCTTAAGATAT 336

seq_name: gb_est33:AI775248

seq_documentation_block:
LOCUS      AI775248      613 bp      mRNA
DEFINITION EST256348 tomato resistant, Cornell Lycopersicon esculentum CDNA
clone CLER15A13, mRNA sequence.
ACCESSION  AI775248
VERSION    AI775248.1  GI:5273289
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Potatoes; Lycopersicon.
            1 (bases 1 to 613)
            D' Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
            Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y.,
            Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
            Giovannoni, J.J. and Martin, G.B.
            Generation of ESTs from Pseudomonas resistant tomato
            Unpublished (1999)
            On May 18, 1998 this sequence version replaced gi:3137089.
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA

```

Tel: 864 656 4366  
 Fax: 864 656 4293  
 Email: dfrisch@CLEMSON.EDU  
 5 prime sequence.

## FEATURES

source  
 Location/Qualifiers  
 1..613  
 /organism="Lycopersicon esculentum"  
 /cultivar="R11-12 (35S::PTO in Rio Grande x Money Maker)"  
 /db\_xref="taxon:4081"  
 /clone="CLER15A13"  
 /clone\_lib="tomato resistant, Cornell"  
 /tissue\_type="leaf"  
 /dev\_stage="4-week old"  
 /lab\_host="SOLR"  
 /note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; CLER - Tomato Pseudomonas Resistant EST Library.  
 Directionally cloned cDNAs inserted into pBlueScript.  
 SK(-) at 5' end with EcoRI and 3' end with XhoI site."  
 BASE COUNT 181 a 147 c 141 g 143 t 1 others  
 ORIGIN

## alignment\_scores:

Quality:	41.00	Length:	11
Ratio:	4.556	Gaps:	0
Percent Similarity:	81.818	Percent Identity:	81.818

## alignment\_block:

US-08-653-294-16 x AI775248 ..

Align seg 1/1 to: AI775248 from: 1 to: 613

1 TyrArgLeuLeuIleArgArgIleLeuLeuArg 11  
 |||||  
 18 TATCGTCCTCTTGTGCCGCGGTATCTTCTACGC 50

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:29:39 ; Search time 122.56 Seconds  
(without alignments)  
2.319 Million cell updates/sec

Title: US-08-653-294-17  
Perfect score: 58  
Sequence: 1 YRLIRRIALRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	91.4	12	1 R95429	HLA-B2702 84-79-84
2	53	91.4	12	1 W33798	Peptide B2702.84-7
3	53	91.4	12	1 W33799	Immunomodulating d
4	39	67.2	20	1 R92907	HLA-B2702 CTL modu
5	39	67.2	20	1 R95428	HLA-B2702 84-75-84
6	39	67.2	20	1 W33778	Immunomodulating d
7	35	62.1	904	1 W86351	Human DNAX toll-11
8	35	60.3	40	1 W12037	Curvularia verrucu
9	35	60.3	600	1 W12042	Curvularia verrucu
10	35	60.3	640	1 R82249	Chloroperoxidase.
11	34	58.6	20	1 R92909	HLA-B2702 CTL modu
12	34	58.6	20	1 R92908	HLA-B2702 CTL modu
13	34	58.6	20	1 W33791	Peptide B2702.84-7
14	34	58.6	20	1 W33792	Peptide B2702.84-7
15	34	58.6	504	1 W44156	Human neuronal nic
16	34	58.6	3391	1 W06591	Polyprotein of att
17	34	58.6	3391	1 W06590	Polyprotein of DEN
18	33	56.9	18	1 R71429	Human MHC 1 alpha
19	33	56.9	327	1 R77379	Alteromonas naneda
20	33	56.9	2240	1 R67819	Acetyl CoA carboxy
21	32	55.2	20	1 R95430	HLA-B2702 84-75T/7
22	32	55.2	495	1 W61177	Ethylenediamine-N,
23	32	55.2	504	1 W09022	Neuronal nicotinic
24	32	55.2	520	1 W74762	Human secreted pro
25	32	55.2	3165	1 R38889	Sequence encoded b
26	31	53.4	31	1 Y10817	Amino acid sequenc
27	31	53.4	3567	1 R44431	eryA region polype
28	31	53.4	4472	1 R97246	Virulence gene clu
29	30	51.7	103	1 Y13111	Human secreted pro
30	30	51.7	111	1 R81449	Hepatitis GB virus
31	30	51.7	227	1 W75423	T.thermophilus nit
32	30	51.7	286	1 R88917	Modified Incc prot
33	30	51.7	529	1 R73966	Alpha 2 subunit of
34	30	51.7	529	1 W09021	Neuronal nicotinic

35 30 51.7 529 1 W44155 Human neuronal nic  
36 30 51.7 627 1 W09023 Neuronal nicotinic  
37 30 51.7 627 1 W11824 Alpha4 subunit of  
38 30 51.7 627 1 W11825 Alpha4 subunit of  
39 30 51.7 1978 1 W69361 Tetradotoxin-sensi  
40 30 51.7 1988 1 W69362 Tetradotoxin-sensi  
41 29.5 50.9 20 1 R92911 HLA-B2702 CTL modu  
42 29.5 50.9 20 1 W33779 Immunomodulating d  
43 29 50.0 6 1 W47264 Immunomodulatory p  
44 29 50.0 6 1 W47261 Immunomodulatory p  
45 29 50.0 6 1 W33783 Peptide #4 used in

## ALIGNMENTS

## RESULT 1

R95429 ID R95429 standard; peptide; 12 AA.  
AC R95429;  
DT 12-NOV-1996 (first entry)  
DE HLA-B2702 84-79-84 palindrome.  
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN W09513288-AL.  
PD 18-MAY-1995.  
PF 10-NOV-1994; UI2985.  
PR 10-NOV-1993; US-150493.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM,  
DR WFI; 95-194027/25.  
PT Compens. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12; 29pp; English.  
CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. This sequence represents the  
CC HLA-B2702 84-79-84 palindrome. These sequences can be used to isolate  
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane  
CC protein associated with T-cell activation in mammalian T-cells, and is  
CC also immunologically cross reactive with the heat shock protein Hsc70.  
CC p74 is found in a limited number of cell types, but is particularly  
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
CC cell with an amphoteric detergent, and then passed through an affinity  
CC column containing a covalently bound HLA-B2702 palindromic peptide.  
CC Compositions comprising the extracellular fragment of p74 combined with  
CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
CC compounds can be screened for their effect on the cytolytic activity of  
CC T-cells, by combining them with the extracellular portion of p74 and  
CC determining the amount of binding between the candidate compound and p74.  
CC Modulation of CTL activity can be inhibited in a cellular composition  
CC containing T-cells and antigen presenting cells (APCs), by adding to the  
CC mix the extracellular portion of p74, in an amount sufficient to compete  
CC with p74 for the binding of the p74 ligand.  
SQ Sequence 12 AA;

Query Match 91.4%; Score 53; DB 1; Length 12;  
Best Local Similarity 91.7%; Pred. No. 0.00035;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
Db 1 YRLAIRRIALRY 12

RESULT 2  
W33798 ID W33798 standard; peptide; 12 AA.  
AC W33798;

DT 19-JUN-1998 (first entry)  
 DE Peptide B2702.84-79/79-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 91.4%; Score 53; DB 1; Length 12;  
 Best Local Similarity 91.7%; Pred. No. 0.00035;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
 ||| |||||  
 Db 1 YRLAIRRIALRY 12

RESULT 3  
 W33799 ID W33799 standard; peptide; 12 AA.  
 AC W33799; 1998 (first entry)  
 DE Immunomodulating dimer peptide #3.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Claim 17; Page 35; 41pp; English.  
 CC This sequence represents a specifically claimed immunomodulating  
 CC dimer peptide of the invention. A peptide-type compound or variant is

CC claimed which has immunomodulating activity, including the N-terminal  
 CC acylated and/or C-terminal amidated or esterified forms of up to 60  
 CC amino acids, where the peptide-type compound comprises the formula: A-B,  
 CC where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or  
 CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a  
 CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa  
 CC represents amino acid. The sequence in the brackets may optionally be  
 CC absent or truncated at any peptide type bond within the brackets. The  
 CC compounds comprise amino acid sequences related to a Class I HLA-B  
 CC alpha domain (positions 79-84). They can be used to inhibit cytotoxic  
 CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in  
 CC vitro. They can also be used in combination with antigenic peptides or  
 CC proteins of interest to activate CTLs. They can also inhibit the  
 CC proliferation of T cells in response to anti-CD3. The peptide can be  
 CC used for preventing rejection of transplants or for treating autoimmune  
 CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.  
 CC The products can also be used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 91.4%; Score 53; DB 1; Length 12;  
 Best Local Similarity 91.7%; Pred. No. 0.00035;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
 ||| |||||  
 Db 1 YRLAIRRIALRY 12

RESULT 4  
 R92907 ID R92907 standard; peptide; 20 AA.

AC R92907;  
 DE 16-MAY-1996 (first entry)  
 DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW Class I MHC; HLA-B2702.  
 OS Synthetic.  
 PN WO9526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI: 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 15; Page 36; 80pp; English.  
 CC R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
 CC I MHC HLA-B2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with a  
 CC subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 SQ Sequence 20 AA;

Query Match 67.2%; Score 39; DB 1; Length 20;  
 Best Local Similarity 55.0%; Pred. No. 0.25;  
 Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 YRLIRRIALRY 12  
 ||| ||  
 Db 1 YRLAIRRIALRY 20  
 |||||



RESULT 5

R95428 ID R95428 standard; peptide; 20 AA.

AC R95428; (first entry)

DT 12-NOV-1996

DE HLA-B\*2702 84-75-84 palindromic.

KW HLA: p74; alpha-helix; human-leucocyte-associated antigen; inhibitor;

KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;

KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;

KW cytolysis; antigen presenting cell.

OS Synthetic.

PN W09513288-A1.

PD 18-MAY-1995.

PF 10-NOV-1994; US-150493.

PR 10-NOV-1993; US-150493.

PA (STRD) UNIV LELAND STANFORD JUNIOR.

PI Clayberger C, Krensky AM;

DR WPI: 95-194027/25.

PT Compsns. comprising lymphoid surface membrane proteins - which may inhibit cytolytic activity and differentiation of CTLs.

PS Example; Page 12; 29pp; English.

CC R95413, and R95415-R95431 represent palindromes and fragments of human-leucocyte-associated antigens. This sequence represents the HLA-B\*2702 84-75-84 palindromic. These sequences can be used to isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane protein associated with T-cell activation in mammalian T-cells, and is also immunologically cross reactive with the heat shock protein Hsc70. p74 is found in a limited number of cell types, but is particularly expressed on B and T cells. p74 can be isolated by lysis of a suitable cell with an amphoteric detergent, and then passed through an affinity column containing a covalently bound HLA-B\*2702 palindromic peptide. Compositions comprising the extracellular fragment of p74 combined with HLA-B\*2702 60-84 (see R95416), induces calcium influx, and inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate compounds can be screened for their effect on the cytolytic activity of T-cells, by combining them with the extracellular portion of p74 and determining the amount of binding between the candidate compound and p74. Modulation of CTL activity can be inhibited in a cellular composition containing T-cells and antigen presenting cells (APCs), by adding to the mix the extracellular portion of p74, in an amount sufficient to compete with p74 for the binding of the p74 ligand.

SQ Sequence 20 AA;

Query Match 67.2%; Score 39; DB 1; Length 20;  
Best Local Similarity 55.0%; Pred. No. 0.25;  
Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Oy 1 YRLLR-----RIALRY 12  
||| |||  
Db 1 YRLAIRLNRERLRIALRY 20

RESULT 6

W33778 ID W33778 standard; peptide; 20 AA.

AC W33778; (first entry)

DT 19-JUN-1998

DE Immunomodulating dimer peptide #1.

KW Immunomodulating dimer; immunosuppressant drug; CTL activation;

KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;

KW rejection.

OS Synthetic.

OS Homo sapiens.

PN W09744351-A1.

PD 27-NOV-1997.

PF 24-MAY-1996; US-653294.

PR (STRD) UNIV LELAND STANFORD JUNIOR.

PA Beulow R, Clayberger C, Krensky AM;

DR WPI: 98-086550/08.

PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B alpha-1 domain, used for preventing rejection of transplants or

treating autoimmune diseases

Claim 16; Page 35; 41pp; English.

This sequence represents a specifically claimed immunomodulating dimer peptide of the invention. A peptide-type compound or variant is claimed which has immunomodulating activity, including the N-terminal acylated and/or C-terminal amidated or esterified forms of up to 60 amino acids, where the peptide-type compound comprises the formula: A-B, where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino acid. The sequence in the brackets may optionally be absent or truncated at any peptide type bond within the brackets. The compounds comprise amino acid sequences related to a Class I HLA-B alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from undesirably attacking cells in a host or in vitro. They can also be used in combination with antigenic peptides or proteins of interest to activate CTLs. They can also inhibit the proliferation of T cells in response to anti-CD3. The peptide can be used for preventing rejection of transplants or for treating autoimmune diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus. The products can also be used for detection and diagnosis.

SQ Sequence 20 AA;

Query Match 67.2%; Score 39; DB 1; Length 20;  
Best Local Similarity 55.0%; Pred. No. 0.25;  
Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Oy 1 YRLLR-----RIALRY 12  
||| |||  
Db 1 YRLAIRLNRERLRIALRY 20

RESULT 7

W86351 ID W86351 standard; Protein; 904 AA.

AC W86351;

DT 15-MAR-1999 (first entry)

DE Human DNAX toll-like receptor DTLR3.

KW DNAX toll-like receptor; DTLR; Drosophila toll receptor; IL-1 receptor;

KW interleukin 1 receptor; phosphate metabolism; innate immunity response;

KW modulate inflammatory function; morphological effect;

OS Homo sapiens.

OS Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label- signal

FT Protein 22..904

FT Protein /label- DTLR3

PN W09850547-A2.

PD 12-NOV-1998.

PF 07-MAY-1998; U08979.

PR 05-MAR-1998; US-076947.

PR 07-MAY-1997; US-044293.

PR 22-JAN-1998; US-072212.

PA (SCHE) SCHERING CORP.

PI Bazan JF, Hardman GT, Kastelein RA, Rock FL;

DR N-PSDB: V80665.

DR Human DNAX toll-like receptors, DTLR 2-10 - used to, e.g. alter phosphate metabolism, modulate inflammatory function or innate immunity responses

PS Claim 2; Page 108-111; 171pp; English.

CC The present invention specifically describes human DNAX toll-like receptors 2 to 10 (DTLR2-10). The present sequence is human DTLR3 given in the present invention. Also described are: (1) a fusion protein comprising a DTLR protein or peptide; (2) a binding compound, preferably an antibody or antibody fragment which specifically binds to a DTLR protein or peptide; (3) a nucleic acid encoding a DTLR protein or peptide; (4) an expression vector comprising the nucleic acid of (3); and (5) a host cell comprising the vector of (4). The host cell of (5) can be used to produce the DTLR proteins. The DTLR proteins can be used to alter phosphate metabolism, to modulate inflammatory function, innate

CC immunity responses or morphological effects. The DTLR proteins can be  
 CC used in the treatment of conditions exhibiting abnormal expression of  
 CC the receptors of their ligands. These abnormalities are typically  
 CC manifested by immunological disorders.  
 SQ Sequence 904 AA;

Query Match 62.1%; Score 36; DB 1; Length 904;  
 Best Local Similarity 60.0%; Pred. No. 44;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLIRRIALR 11  
 |::|::|::|  
 Db 484 RLMLRRVALK 493

## RESULT 8

WI2037  
 ID WI2037 standard; Peptide; 40 AA.  
 AC WI2037;  
 DT 30-APR-1997 (first entry)  
 DE Curvularia verruculosa haloperoxidase partial peptide (aa275-314).  
 KW Haloperoxidase; halide oxidation; halogenation; antimicrobial;  
 KW disinfectant.  
 OS Curvularia verruculosa strain CBS 147.63.  
 PN WO9704102-A1.  
 PD 06-FEB-1997.  
 PF 09-JUL-1996; U11458.  
 PR 14-JUL-1995; US-001194.  
 PR 21-FEB-1996; US-603534.  
 PA (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO-NORDISK AS.  
 PI Berka RM, Cherry J, Fuglsang C, Halkier T, Oxenboll KM;  
 WPI; 97-132641/12.  
 DR Haloperoxidase from Curvularia verruculosa - useful for oxidn. of  
 PT halide, halogenation and, in presence of hydrogen peroxide and  
 PT thiocyanate, as antimicrobial  
 PS Claim 17; Page 38; 58pp; English.  
 CC A haloperoxidase partial peptide (WI2037) corresponds to amino  
 CC acids 275-314 of the Curvularia verruculosa CBS 63 enzyme. Partial  
 CC peptides (WI2034-41) were obtd. from the haloperoxidase by  
 CC digestion with lysyl-specific protease from Achromobacter. Primers  
 CC (see also T51613-14) based on the peptides were used in the PCR  
 CC amplification of genomic DNA from CBS 147.63. The PCR product was  
 CC used to screen a genomic library, leading to the isolation of the  
 CC haloperoxidase gene (T51612).  
 SQ Sequence 40 AA;

Query Match 60.3%; Score 35; DB 1; Length 40;  
 Best Local Similarity 50.0%; Pred. No. 2.8;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
 |::|::|::|  
 Db 27 YNQIVRRIVATY 38

## RESULT 9

WI2042  
 ID WI2042 standard; Protein; 600 AA.  
 AC WI2042;  
 DT 30-APR-1997 (first entry)  
 DE Curvularia verruculosa haloperoxidase.  
 KW Haloperoxidase; halide oxidation; halogenation; antimicrobial;  
 KW disinfectant.  
 OS Curvularia verruculosa strain CBS 147.63.  
 PN WO9704102-A1.  
 PD 06-FEB-1997.  
 PF 09-JUL-1996; U11458.  
 PR 14-JUL-1995; US-001194.  
 PR 21-FEB-1996; US-603534.  
 PA (NOVO ) NOVO NORDISK BIOTECH INC.

PA (NOVO ) NOVO-NORDISK AS.  
 PI Berka RM, Cherry J, Fuglsang C, Halkier T, Oxenboll KM;  
 WPI; 97-132641/12.  
 DR N-PSDB; T51612.  
 PT Haloperoxidase from Curvularia verruculosa - useful for oxidn. of  
 PT thiocyanate, as antimicrobial  
 PS Claim 23; Page 36-37; 58pp; English.  
 CC The haloperoxidase of Curvularia verruculosa CBS.63 shows optimum  
 CC activity at about 60 deg C and pH 5.5; retains at least 75%  
 CC activity after incubation for 1 hr at pH 7.0 and 60 deg C in the  
 CC presence of 0.1% H2O2, and prefers bromide over chloride ions as  
 CC substrate. It can be obtd. from a C. verruculosa fermentation  
 CC or by expression in host cells transformed with a vector carrying  
 CC the haloperoxidase gene (see also T51612). The isolated enzyme can  
 CC be used to oxidise halide to hypohalous acid in the presence of an  
 CC H2O2 source, or to halogenate cpds. It can also, in the presence  
 CC of H2O2 and thiocyanate, be used as an antimicrobial.  
 SQ Sequence 600 AA;

Query Match 60.3%; Score 35; DB 1; Length 600;  
 Best Local Similarity 50.0%; Pred. No. 44;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
 |::|::|::|  
 Db 301 YNQIVRRIVATY 312

## RESULT 10

R82249  
 ID R82249 standard; Protein; 640 AA.  
 AC R82249;  
 DT 25-APR-1996 (first entry)  
 DE Chloroperoxidase.  
 KW Chloroperoxidase; haloperoxidase; antifouling paint; preservative;  
 KW halide; chloride.  
 OS Curvularia inaequalis.  
 PN WO9527009-A1.  
 PD 12-OCT-1995.  
 PF 30-MAR-1995; NL0123.  
 PR 31-MAR-1994; EP-200893.  
 PR 24-JUN-1994; NL-001048.  
 PA (SCHE-) STICHTING SCHEIKUNDIG ONDERZOEK IN NEDER.  
 PA (TENE-) STICHTING TECH.  
 PA WETENSCHAPPEN.  
 PI Dekker HL, Van Schijndel JWP, Vollenbroek ECM, Wever R;  
 WPI; 95-358602/46.  
 DR N-PSDB; T03875.  
 PT Use of halo:peroxidase enzymes - in antifouling paints and methods  
 PT for determining the halide concn. of liquids  
 PS Example 3; Fig 6; 49pp; English.  
 CC A chloroperoxidase (R82249) is obtd. by expression of a cDNA sequence  
 CC (T03875) isolated from Curvularia inaequalis in a procarvotic host.  
 CC The apoprotein is activated by addn. of vanadate. The enzyme displays  
 CC high thermostability (Tm 90 deg) and is stable in 40% methanol, ethanol  
 CC or propanol. It can be used for the analysis of halide (esp. chloride)  
 CC ions, e.g. in blood or urine. The enzyme may also be utilised as an  
 CC environmentally-friendly antifouling agent in ship paint; it  
 CC oxidises bromide ions present in seawater to form bactericidal HOBr.  
 SQ Sequence 640 AA;

Query Match 60.3%; Score 35; DB 1; Length 640;  
 Best Local Similarity 50.0%; Pred. No. 47;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
 |::|::|::|  
 Db 327 YNQIVRRIVATY 338

```

RESULT 11
R92909
ID R92909 standard; peptide; 20 AA.
AC R92909;
DT 16-MAY-1996 (first entry)
DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLA-B2702.
OS Synthetic.
PN W09526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.
PR 05-APR-1994; US-222851.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI; 95-358582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B75-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC I MHC HLA-B2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RIALRY 12
   ||| | | | | |
Db 1 YRLAIRNERENRIALRY 20

RESULT 12
R92908
ID R92908 standard; peptide; 20 AA.
AC R92908;
DT 16-MAY-1996 (first entry)
DE HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLA-B2702.
OS Synthetic.
PN W09526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.
PR 05-APR-1994; US-222851.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI; 95-358582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B75-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC I MHC HLA-B2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RIALRY 12
   ||| | | | | |
Db 1 YRLAIRNERENRIALRY 20

RESULT 13
W33791
ID W33791 standard; peptide; 20 AA.
AC W33791;
DT 19-JUN-1998 (first entry)
DE Peptide B2702.84-75/75-84 tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
PN W09744351-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise
CC amino acid sequences related to a Class I HLA-B alpha domain (positions
CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC undesirably attacking cells in a host or in vitro. They can also be
CC used in combination with antigenic peptides or proteins of interest to
CC activate CTLs. They can also inhibit the proliferation of T cells in
CC response to anti-CD3. The peptide can be used for preventing rejection
CC of transplants or for treating autoimmune diseases, e.g. diabetes,
CC rheumatoid arthritis and lupus erythematosus. The products can also be
CC used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RIALRY 12
   ||| | | | | |
Db 1 YRLAIRNERENRIALRY 20

RESULT 14
W33792
ID W33792 standard; peptide; 20 AA.
AC W33792;
DT 19-JUN-1998 (first entry)

```

```

CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RIALRY 12
   ||| | | | | |
Db 1 YRLAIRNERENRIALRY 20

RESULT 13
W33791
ID W33791 standard; peptide; 20 AA.
AC W33791;
DT 19-JUN-1998 (first entry)
DE Peptide B2702.84-75/75-84 tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
PN W09744351-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise
CC amino acid sequences related to a Class I HLA-B alpha domain (positions
CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC undesirably attacking cells in a host or in vitro. They can also be
CC used in combination with antigenic peptides or proteins of interest to
CC activate CTLs. They can also inhibit the proliferation of T cells in
CC response to anti-CD3. The peptide can be used for preventing rejection
CC of transplants or for treating autoimmune diseases, e.g. diabetes,
CC rheumatoid arthritis and lupus erythematosus. The products can also be
CC used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RIALRY 12
   ||| | | | | |
Db 1 YRLAIRNERENRIALRY 20

RESULT 14
W33792
ID W33792 standard; peptide; 20 AA.
AC W33792;
DT 19-JUN-1998 (first entry)

```

DE Peptide B2702.84-75/75-84T tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08589.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 PT alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1: Page 19; 41pp; English.  
 CC Peptides W3784-98 and W3778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha1 domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;  
 Best Local Similarity 50.0%; Pred. No. 2.1;  
 Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLIR-----RIALRY 12  
 ||| ||  
 1 YRLAIRLNERENLRALRY 20

RESULT 15  
 W44156  
 ID W44156 standard; Protein: 504 AA.

AC W44156;  
 DT 14-MAY-1998 (first entry)  
 DE Human neuronal nicotinic acetylcholine receptor alpha-3 subunit.  
 KW Human; neuronal nicotinic acetylcholine receptor; alpha-3 subunit;  
 KW brain tissue; screening; NAChr; antibody.  
 OS Homo sapiens.

PH Key Location/Qualifiers  
 FT Peptide 1..30  
 FT /label= signal  
 FT Misc\_difference 235  
 FT /label= unspecified  
 FT /note= "encoded by WWC"  
 FT Domain 240..265  
 FT /label= TMD1  
 FT /note= "transmembrane domain"  
 FT Misc\_difference 246  
 FT /label= unspecified  
 FT /note= "encoded by AWC"  
 FT Domain 273..296  
 FT /label= TMD2  
 FT /note= "transmembrane domain"  
 FT Misc\_difference 275

FT /note= "encoded by TGY"  
 FT 302..326  
 FT /label= TMD3  
 FT /note= "transmembrane domain"  
 FT Misc\_difference 309  
 FT /label= unspecified  
 FT /note= "encoded by WWC"  
 FT Misc\_difference 347  
 FT /label= unspecified  
 FT /note= "encoded by AWC"  
 FT Misc\_difference 354  
 FT /label= unspecified  
 FT /note= "encoded by TWC"  
 FT 459..480  
 FT /label= TMD4  
 FT /note= "transmembrane domain"  
 FT 327..458  
 FT /label= cytoplasmic\_loop

W09420617-A2.  
 15-SEP-1994.  
 08-MAR-1994; U02447.  
 08-MAR-1993; US-028031.  
 (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Elliott KJ, Ellis SB, Harpold MM;  
 DR WPI: 94-303024/37.  
 DR N-PSDB: V12200.  
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
 PT also transformed cells useful for screening cpds. which modulate  
 PT activity of the receptor  
 PS Example 2: Page 72-73; 99pp; English.  
 CC The present sequence represents a human neuronal nicotinic acetylcholine  
 CC receptor (NAChr) subunit. The cells expressing the alpha and/or beta  
 CC NAChr subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal NAChr.  
 CC Subunit specific density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits or specific receptor  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes.  
 SQ Sequence 504 AA;

Query Match 58.6%; Score 34; DB 1; Length 504;  
 Best Local Similarity 58.3%; Pred. No. 57;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12  
 ||| |||::||  
 Db 232 YSLXIRLSLFY 243

Search completed: February 8, 2000, 01:29:40  
 Job time: 1752 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 11:54:27 ; Search time 117.7 Seconds  
(without alignments)  
4.809 Million cell updates/sec

Title: US-08-653-294-17

Perfect score: 58

Sequence: 1 YRLIRIALRY 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR\_62.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	63.8	313	2 JC5342	Na+/H+ antiporter
2	35	60.3	309	2 G70882	probable oxidoreductase
3	35	60.3	609	2 S69334	chloride peroxidase
4	34	58.6	88	2 C70200	ribosomal protein
5	34	58.6	89	2 H70883	probable ribosomal
6	34	58.6	362	2 G71167	hypothetical prote
7	34	58.6	502	1 OXCKAX	acyl-CoA oxidase (
8	34	58.6	709	1 OXCKX4	acyl-CoA oxidase (
9	34	58.6	709	1 OXCKX	acyl-CoA oxidase (
10	34	58.6	709	1 OXCKPM	acyl-CoA oxidase (
11	34	58.6	3391	1 GNMV16	genome polyprotein
12	34	58.6	3391	1 GNMV26	conserved hypotet
13	33	56.9	107	1 E69263	hypothetical prote
14	33	56.9	352	2 S38147	hypothetical prote
15	33	56.9	550	2 E72481	hypothetical prote
16	33	56.9	869	2 S49844	probable membrane
17	33	56.9	2304	2 T07920	probable acetyl-Co
18	32	55.2	78	2 F64429	DNA-directed RNA p
19	32	55.2	157	2 H75054	hypothetical prote
20	32	55.2	185	2 JC4867	TB2/DPI protein ho
21	32	55.2	288	2 D71325	conserved hypotet
22	32	55.2	298	2 S53849	ribosomal protein
23	32	55.2	310	2 A33489	hypothetical prote
24	32	55.2	399	1 S36553	kpsD protein - Esc
25	32	55.2	408	2 D70365	fimbrial assembly
26	32	55.2	416	1 ACCB3N	nicotinic acetylch
27	32	55.2	469	2 F71307	hypothetical prote
28	32	55.2	470	2 G70564	probable membrane
29	32	55.2	495	2 S60589	acetylcholine rece
30	32	55.2	499	2 A24572	nicotinic acetylch

31 32 55.2 503 2 A53956 nicotinic acetylch  
32 32 55.2 619 2 G72709 probable DNA ligas  
33 32 55.2 636 2 C70031 hypothetical prote  
34 32 55.2 895 2 D72571 probable DNA-direc  
35 32 55.2 1998 2 T08822 nonstructural poly  
36 32 55.2 3165 2 S15010 hypothetical prote  
37 31 53.4 89 2 S74731 ribosomal protein  
38 31 53.4 238 2 S42124 LiCA protein - Myc  
39 31 53.4 246 2 JC1496 transcription fact  
40 31 53.4 272 2 A70683 probable cysw - Wy  
41 31 53.4 306 2 F72323 conserved hypotet  
42 31 53.4 334 2 H64566 ferrocyclase (EC  
43 31 53.4 335 2 A71860 ferrocyclase - H  
44 31 53.4 348 2 S29990 histocompatibility  
45 31 53.4 361 2 A75148 lsu ribosomal prot

## ALIGNMENTS

RESULT 1  
JC5342  
Na+/H+ antiporter cdv2 - Clostridium difficile  
C:Species: Clostridium difficile  
C:Date: 27-May-1997 #sequence\_revision 18-Jul-1997 #text\_change 10-Sep-1997  
C:Accession: JC5342  
R:Braun, V.; Hundsberger, T.; Leukel, P.; Sauerborn, M.; von Eichel-Streiber, C.  
Gene 181, 29-38, 1996  
A:Title: Definition of the single integration site of the pathogenicity locus in Cios  
A:Reference number: JC5340; MUID:97128764  
A:Accession: JC5342  
A:Molecule type: DNA  
A:Residues: 1-313 <BRA>  
A:Cross-references: EMBL:X92982; NID:g1770128; PID:e212007; PID:g1770130  
A:Experimental source: strain VP10463  
C:Comment: This protein contains 11 membrane spanning domain. It functionally links t  
C:Genetics:  
A:Gene: cdv2

Query Match 63.8%; Score 37; DB 2: Length 313;  
Best Local Similarity 41.7%; Pred. No. 8;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLIRIALRY 12  
|||:|:|:|:|  
Db 259 YRVARRLSIKY 270

RESULT 2  
G70882  
Probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 29-Sep-1999  
C:Accession: G70882  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Nature 393, 537-544, 1998  
A:Authors: Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Qua  
; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: G70882  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-309 <COL>  
A:Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA15591.1; PID:e129  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv2776c  
C:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; fer  
F:9-211/Domain: cytochrome-b5 reductase homology <CBR>  
F:240-297/Domain: ferredoxin [2Fe-2S] homology <FER>

Query Match 60.3%; Score 35; DB 2; Length 309;

Best Local Similarity 77.8%; Pred. No. 19;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLIRRIA 9

II : IIII

DB 65 YRIAIRIA 73

RESULT 3

S69334

chloride peroxidase (EC 1.11.1.10) - Curvularia inaequalis

N:Alternate names: vanadium-containing chloroperoxidase

C:Species: Curvularia inaequalis

C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999

R:Accession: S69334; S53117

R:Simons, B.H.; Barnett, P.; Vollenbroek, E.G.M.; Dekker, H.L.; Mulijers, A.O.; Messers

Eur. J. Biochem. 229, 566-574, 1995

A:Title: Primary structure and characterization of the vanadium chloroperoxidase from th

A:Reference number: S69334; MUID:95262722

A:Accession: S69334

A:Molecule type: mRNA

A:Residues: 1-609 <SIM>

A:Cross-references: EMBL:X85369; NID:g732617; PID:g732618

A:Experimental source: CBS 102.42

A>Note: part of this sequence was confirmed by sequence analysis of genomic DNA

R:Messerschmidt, A.; Wever, R.

submitted to the Brookhaven Protein Data Bank, September 1995

A:Reference number: A66871; PDB:1VNC

A:Contents: annotation: X-ray crystallography, 2.1 angstroms, 3-543, 'E', 545-578

A>Note: the authors believe that 544-Glu is correct from modeling evidence

C:Genetics:

A:Gene: VCPO

C:Function:

A:Description: catalyzes the formation of two carbon-chlorine bonds and two water molecu

C:Superfamily: Curvularia inaequalis chloride peroxidase

C:Keywords: blocked amino end; chloride; metalloprotein; oxidoreductase; vanadium

F:404/Active site: His #status predicted

F:496/Binding site: vanadate (His) #status predicted

Query Match 60.3%; Score 35; DB 2; Length 609;

Best Local Similarity 50.0%; Pred. No. 38;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12

II : IIII

DB 304 YNQIVRIATVY 315

RESULT 4

C70200

ribosomal protein S15 (rpsO) - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 13-Aug-1999

C:Accession: C70200

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Kervilavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943

A:Accession: C70200

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-88 <GLE>

A:Cross-references: GB:AE001179; GB:AE000783; NID:g2688738; PIDN:AA067150.1; PID:g268874

A:Experimental source: strain B31

C:Superfamily: Escherichia coli ribosomal protein S15; eubacterial ribosomal protein S15

F:22-88/Domain: eubacterial ribosomal protein S15 homology <ES15>

Query Match 58.6%; Score 34; DB 2; Length 88;

Best Local Similarity 54.5%; Pred. No. 8.9;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLIRRIALR 11

II : III : II

DB 77 YRLISLGLR 87

RESULT 5

H70883

probable ribosomal protein S15 rpsO - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Sep-1999

C:Accession: H70883

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: H70883

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-89 <COL>

A:Cross-references: GB:AL123456; GB:AL123456; NID:g3261491; PIDN:CAAL580.1; PID:e117

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: rpsO

C:Superfamily: Escherichia coli ribosomal protein S15; eubacterial ribosomal protein

F:23-89/Domain: eubacterial ribosomal protein S15 homology <ES15>

Query Match 58.6%; Score 34; DB 2; Length 89;

Best Local Similarity 63.6%; Pred. No. 9;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALR 11

II : III : II

DB 78 YRLIRLGLR 88

RESULT 6

G71167

hypothetical protein PH0539 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998

C:Accession: G71167

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: G71167

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-362 <RAW>

A:Cross-references: GB:AP000002; NID:g3236129; PID:d1030571; PID:g3256945

A:Experimental source: strain OT3

A>Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0539

Query Match 58.6%; Score 34; DB 2; Length 362;

Best Local Similarity 50.0%; Pred. No. 35;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALRY 12

II : III : II

Db 236 YRFLKATG 307

RESULT 7

OXCKAX

acyl-CoA oxidase (EC 1.3.3.6) POX4-2, peroxisomal - yeast (*Candida tropicalis*) (fragment)  
N:Alternate names: acyl-CoA oxidase II-2  
C:Species: *Candida tropicalis*  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 11-Jun-1999  
C:Accession: A28584  
R:Small, G.W.; Lazarow, P.B.  
J. Cell Biol. 105, 247-250, 1987  
A:Title: Import of the carboxy-terminal portion of acyl-CoA oxidase into peroxisomes of  
A:Reference number: A28584; MUID:87280361  
A:Molecule type: mRNA  
A:Accession: A28584  
A:Residues: 1-502 <SMA>  
A:Cross-references: GB:Y00623; NID:g2672; PIDN:CAA68660.1; PID:g2673  
C:Genetics:  
A:Gene: POX4-2  
C:Superfamily: acyl-CoA oxidase  
C:Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match 58.6%; Score 34; DB 1; Length 502;

Best Local Similarity 64.3%; Pred. No. 49;

Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 YRLLR--RIALRY 12

|||||

Db 101 YRLARMSTIALRY 114

RESULT 8

OXCK4

acyl-CoA oxidase (EC 1.3.3.6) POX4, peroxisomal - yeast (*Candida tropicalis*)  
N:Alternate names: acyl-CoA oxidase II  
C:Species: *Candida tropicalis*  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 11-Jun-1999  
C:Accession: A25123  
R:Okazaki, K.; Takechi, T.; Kanbara, N.; Fukui, S.; Kubota, I.; Kamiryo, T.  
Proc. Natl. Acad. Sci. U.S.A. 83, 1232-1236, 1986  
A:Title: Two acyl-coenzyme A oxidases in peroxisomes of the yeast *Candida tropicalis*: pr  
A:Reference number: A94084; MUID:86149279  
A:Accession: A25123  
A:Molecule type: DNA  
A:Residues: 1-709 <ONA>  
A:Cross-references: GB:M12160; NID:g170911; PIDN:AAA34362.1; PID:g170912  
A:Experimental source: strain pK233, ATCC 20336  
C:Comment: This enzyme, located in peroxisomes, catalyzes the oxygen-specific oxidation  
; this reaction is the initial step of the peroxisomal beta-oxidation system. In *C. trop*  
gene family.  
C:Comment: The peroxisomes are eukaryote subcellular organelles that generate and degrad  
ic reactions as well as the enzymes involved are usually different from the mitochondria  
C:Genetics:  
A:Gene: POX4  
C:Superfamily: acyl-CoA oxidase  
C:Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match 58.6%; Score 34; DB 1; Length 709;

Best Local Similarity 64.3%; Pred. No. 68;

Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 YRLLR--RIALRY 12

|||||

Db 308 YRLARMSTIALRY 321

RESULT 9

OXCKX

acyl-CoA oxidase (EC 1.3.3.6) AOX, peroxisomal - yeast (*Candida tropicalis*)  
C:Species: *Candida tropicalis*

C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 30-Sep-1993  
C:Accession: A29047  
R:Murray, W.W.; Rachubinski, R.A.  
Gene 51, 119-128, 1987

A:Title: The primary structure of a peroxisomal fatty acyl-CoA oxidase from the yeast  
A:Reference number: A29047; MUID:87248070  
A:Accession: A29047

A:Molecule type: DNA

A:Residues: 1-709 <MUR>

A:Experimental source: strain pK233, ATCC 20336

C:Genetics:

A:Gene: AOX

C:Superfamily: acyl-CoA oxidase

C:Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match 58.6%; Score 34; DB 1; Length 709;

Best Local Similarity 64.3%; Pred. No. 68;

Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 YRLLR--RIALRY 12

|||||

Db 308 YRLARMSTIALRY 321

RESULT 10

OXCKPM

acyl-CoA oxidase (EC 1.3.3.6) PXP4, peroxisomal - yeast (*Candida maltosa*)

C:Species: *Candida maltosa*

C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 17-Feb-1995

C:Accession: A29441

R:Hill, D.E.; Boulay, R.; Rogers, D.

Nucleic Acids Res. 16, 365-366, 1988

A:Title: Complete nucleotide sequence of the peroxisomal acyl CoA oxidase from the al

A:Reference number: A29441; MUID:88124223

A:Accession: A29441

A:Molecule type: DNA

A:Residues: 1-709 <HIL>

A:Experimental source: ATCC 20184

C:Genetics:

A:Gene: POX4

C:Superfamily: acyl-CoA oxidase

C:Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match 58.6%; Score 34; DB 1; Length 709;

Best Local Similarity 64.3%; Pred. No. 68;

Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 YRLLR--RIALRY 12

|||||

Db 308 YRLARMSTIALRY 321

RESULT 11

GNMV16

genome polyprotein - dengue virus type 2 (strain 16681)  
N:Contains: capsid protein C; envelope protein E; membrane-associated protein M; nons  
tural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: dengue virus type 2

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 16-Jul-1999

C:Accession: A42451; A43496; A43763

R:Blot, J.; McWilliam, S.M.; Butler, H.C.; Gibbs, A.J.; Weiller, G.; Herring, B.L.; H

Virology 187, 573-590, 1992

A:Title: Comparison of a dengue-2 virus and its candidate vaccine derivative: sequenc

A:Reference number: A42451; MUID:92188532

A:Accession: A42451

A:Molecule type: genomic RNA

A:Residues: 1-3391 <BLO>

A:Cross-references: GB:M84727; GB:M85259; NID:g323472; PIDN:AAA73185.1; PID:g323473

C:Superfamily: yellow fever virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protei

F:1-114/Product: capsid protein C #status predicted <CPC>

F:115-280/Product: membrane-associated protein M precursor #status predicted <MPP>  
 F:115-205/Domain: nonterminal signal sequence #status predicted <SIG>  
 F:206-280/Product: membrane-associated protein M #status predicted <MPM>  
 F:268-284/Domain: transmembrane #status predicted <TM1>  
 F:281-775/Product: envelope protein E #status predicted <EPE>  
 F:727-773/Domain: transmembrane #status predicted <TM2>  
 F:757-773/Domain: transmembrane #status predicted <TM3>  
 F:1128-1345/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1128-1474/Product: nonstructural protein NS2a #status predicted <NS2a>  
 F:1346-1474/Product: nonstructural protein NS2b #status predicted <NS2b>  
 F:1475-2093/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1668-1938/Domain: DEAD/H box helicase homology <DEAD>  
 F:1668-1675/Region: nucleotide-binding motif A (P-loop)  
 F:1755-1760/Region: nucleotide-binding motif A (P-loop)  
 F:1759-1762/Region: DEAH motif  
 F:2094-2243/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:2244-2491/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2492-3391/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:183-347,433/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 58.6%; Score 34; DB 1; Length 3391;

Best Local Similarity 54.5%; Pred. No. 3.2e+02;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLIRIRIALRY 12

DB 2728 RMLINRFTMY 2738

#### RESULT 12

GNVW26

N:Contains: capsid protein C; envelope protein E; membrane-associated protein M; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: dengue virus type 2  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 15-May-1998  
 C:Accession: B42451

A:Title: Comparison of a dengue-2 virus and its candidate vaccine derivative: sequence 1  
 A:Reference number: A42451; MUID:92188532  
 A:Accession: B42451

A:Molecule type: genomic RNA

A:Residues: 1-3391 <BLQ>

A:Cross-references: GB:M85259

C:Superfamily: yellow fever virus genome polyprotein; DEAD/H box helicase homology  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;  
 F:1114/Product: capsid protein C #status predicted <CPC>  
 F:50-66/Domain: transmembrane #status predicted <TM1>

F:102-118/Domain: transmembrane #status predicted <TM2>

F:115-280/Product: membrane-associated protein M precursor #status predicted <MPP>

F:115-205/Domain: nonterminal signal sequence #status predicted <SIG>

F:206-280/Product: membrane-associated protein M #status predicted <MPM>

F:268-284/Domain: transmembrane #status predicted <TM3>

F:281-775/Product: envelope protein E #status predicted <EPE>

F:727-743/Domain: transmembrane #status predicted <TM4>

F:757-773/Domain: transmembrane #status predicted <TM5>

F:776-1127/Product: nonstructural protein NS1 #status predicted <NS1>

F:1128-1345/Product: nonstructural protein NS2a #status predicted <NS2a>

F:1158-1174/Domain: transmembrane #status predicted <TM6>

F:1272-1288/Domain: transmembrane #status predicted <TM7>

F:1294-1310/Domain: transmembrane #status predicted <TM8>

F:1346-1474/Product: nonstructural protein NS2b #status predicted <NS2b>

F:1351-1367/Domain: transmembrane #status predicted <TM9>

F:1373-1389/Domain: transmembrane #status predicted <TM6>

F:1448-1464/Domain: transmembrane #status predicted <TM6>

F:1475-2093/Product: nonstructural protein NS3 #status predicted <NS3>

F:1668-1938/Domain: DEAD/H box helicase homology <DEAD>

F:1668-1675/Region: nucleotide-binding motif A (P-loop)

F:1755-1760/Region: nucleotide-binding motif A (P-loop)

F:1759-1762/Region: DEAH motif

F:2094-2243/Product: nonstructural protein NS4a #status predicted <NS4a>

F:2148-2164/Domain: transmembrane #status predicted <TMC>  
 F:2174-2190/Domain: transmembrane #status predicted <TMD>  
 F:2197-2213/Domain: transmembrane #status predicted <TME>  
 F:2227-2243/Domain: transmembrane #status predicted <TMF>  
 F:2244-2491/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2352-2368/Domain: transmembrane #status predicted <TMG>  
 F:2411-2427/Domain: transmembrane #status predicted <TMH>  
 F:2492-3391/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:183-347,433,905,982,1134,1174,1329,2301,2305,2346,2387,2457,2485,2644,2665,2704,271

Query Match 58.6%; Score 34; DB 1; Length 3391;

Best Local Similarity 54.5%; Pred. No. 3.2e+02;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLIRIRIALRY 12

DB 2728 RMLINRFTMY 2738

#### RESULT 13

E69263

Conserved hypothetical protein AF0109 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: E69263

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343

A:Accession: E69263

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-107 <KLE>

A:Cross-references: GB:AF001099; GB:AF000782; NID:g2689422; PID:g2650539; TIGR:AF0109

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0962

Query Match 56.9%; Score 33; DB 1; Length 107;

Best Local Similarity 58.3%; Pred. No. 17;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YLLIRIRIALRY 12

DB 38 YVELARKIAMRY 49

#### RESULT 14

S38147

Hypothetical protein YKR070w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 14-Nov-1997

C:Accession: S38147

R:Pohl, T.M.; Pohl, F.M.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S37897

A:Accession: S38147

A:Molecule type: DNA

A:Residues: 1-352 <POH>

A:Cross-references: EMBL:Z28295; NID:g486540; PID:g486541; MIPS:YKR070w

A:Experimental source: strain S28C

C:Genetics:

A:Map position: 11R

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Score

Pred. No.



QY 1 YRLIRRIAL 10  
: : : : :  
Db 223 FRLVRRRLYL 232

## RESULT 15

E72481  
hypothetical protein APE2493 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: E72481  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy  
A:Reference number: A72450; MUID:99310339  
A:Accession: E72481  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-550 <RAW>  
A:Cross-references: DDBJ:AP000064; NID:95105945; PIDN:BAA81509.1; PID:d1045295; PID:g510  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2493

Query Match 56.9%; Score 33; DB 2; Length 550;  
Best Local Similarity 60.0%; Pred. No. 84;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLLIRRIALR 11  
: : : : :  
Db 405 RLLVRRVNR 414

Search completed: February 7, 2000, 11:54:28  
Job time: 24338 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 00:59:55 ; Search time 63.71 seconds  
(without alignments)  
5.625 Million cell updates/sec

Title: US-08-653-294-17  
Perfect score: 58  
Sequence: 1 YRLIRRIALRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	35.5	61.2	491	1 HEMN_ALCEU	O34162 alcaligenes
2	35	60.3	609	1 PRXC_CURIN	P49053 curvularia
3	34	58.6	88	1 RS15_BORBU	O51744 borrelia bu
4	34	58.6	89	1 RS15_MYCLE	O32967 mycobacteri
5	34	58.6	89	1 RS15_MYCTU	O33327 mycobacteri
6	34	58.6	411	1 YHS5_CABEL	O18304 caenorhabdi
7	34	58.6	502	1 CAO3_CANTR	P11355 candida tiro
8	34	58.6	708	1 CAO3_CANTR	P65598 candida tiro
9	34	58.6	708	1 CAO4_CANNA	P53335 candida mal
10	34	58.6	3391	1 POLG_DEN26	P29990 d genome po
11	34	58.6	3391	1 POLG_DEN27	P29991 d genome po
12	33	56.9	352	1 YK50_YEAST	P36151 saccharomyc
13	33	56.9	869	1 YD95_YEAST	P38966 saccharomyc
14	32	55.2	78	1 RPOH_METJA	Q38443 methanococc
15	32	55.2	185	1 DFL_MOUSE	Q60870 mus musculu
16	32	55.2	298	1 RT03_ACACA	Q60870 mus musculu
17	32	55.2	389	1 KSS5_ECOLI	P42218 escherichia
18	32	55.2	495	1 ACH3_BOVIN	Q07263 bos taurus
19	32	55.2	496	1 ACH3_CHICK	P09481 gallus gall
20	32	55.2	499	1 ACH3_RAT	P04757 rattus norv
21	32	55.2	503	1 ACH3_HUMAN	P32297 homo sapien
22	31	53.4	89	1 RS15_SYNY3	P72866 synecocyst
23	31	53.4	238	1 LICA_MYCCA	P43044 mycoplasma
24	31	53.4	246	1 MTT1_HUMAN	Q00059 homo sapien
25	31	53.4	334	1 HEM2_HELPY	P56107 helicobacte
26	31	53.4	348	1 HLAF_MACMU	P33617 macaca mula
27	31	53.4	443	1 COXX_HUMAN	Q12887 homo sapien
28	31	53.4	454	1 ACH5_CHICK	P21652 gallus gall
29	31	53.4	456	1 HEMN_AQUAE	O67886 aquifex aeo
30	31	53.4	487	1 ACH6_CABEL	Q93149 caenorhabdi
31	31	53.4	494	1 EXOT_RHINE	P33699 rhizobium m
32	31	53.4	617	1 SIP_TREPA	O83195 treponema p
33	31	53.4	622	1 ACH4_CHICK	P09482 gallus gall
34	31	53.4	633	1 ACH4_RAT	P09483 rattus norv

```

35 31 53.4 1151 1 GRR1_YEAST P24814 saccharomyc
36 31 53.4 2168 1 POLG_PEV9U O41174 p genome po
37 31 53.4 3567 1 ERY2_SACER Q03132 saccharopol
38 30.5 52.6 233 1 PRC3_CARAU O73672 xenopus a
39 30.5 52.6 233 1 PRC3_XENLA P24495 xenopus lae
40 30.5 52.6 234 1 PRC3_DROME P40301 drosophila
41 30.5 52.6 278 1 YD87_YEAST P35178 saccharomyc
42 30.5 52.6 1229 1 SIP3_YEAST P38717 saccharomyc
43 30 51.7 88 1 RS15_BACSU P21473 bacillus su
44 30 51.7 88 1 RS15_THETH P80378 thermus aqu
45 30 51.7 95 1 CH10_ZYMO P48229 zymomonas m

```

## ALIGNMENTS

```

* RESULT 1
ID HEMN_ALCEU STANDARD; PRT; 491 AA.
AC O34162;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE OXYGEN-INDEPENDENT COPROPORPHYRINOGEN III OXIDASE (EC 1.-.-.-)
DE (COPROPORPHYRINOGENASE) (COPROGEN OXIDASE).
GN HEMN.
OS Alcaligenes eutrophus.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Raistonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H16;
RX MEDLINE; 98060717.
RA LIEB C., SIDDIQUI R.A., HIPPLER B., JAHN D., FRIEDRICH B.;
RT "The Alcaligenes eutrophus hemn gene encoding the oxygen-independent
RT coproporphyrinogen III oxidase, is required for heme biosynthesis
RT during anaerobic growth."
RL Arch. Microbiol. 169:52-60(1998).
CC -1- FUNCTION: ANAEROBIC TRANSFORMATION OF COPROPORPHYRINOGEN-III INTO
CC PROTOPORPHYRINOGEN-IX.
CC -1- COFACTOR: REQUIRES MAGNESIUM, ATP AND NAD (OR NADP) FOR ACTIVITY.
CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ANAEROBIC COPROPORPHYRINOGEN III
CC OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U94742; AAB66374.1;
CC Porphyrin biosynthesis; Oxidoreductase; Magnesium; NAD.
CC SEQUENCE 491 AA; 54216 MW; 3EE82375 CRC32;
Query Match 61.2%; Score 35.5; DB 1; Length 491;
Best Local Similarity 57.1%; Pred. No. 11;
Matches 8; Conservative 3; Mismatches 0; Indels 3; Gaps 1;
QY 2 RLLIRRIAL---RY 12
|||:|:|:|
Db 441 RLLVRRVAVFDY 454
RESULT 2
PRXC_CURIN STANDARD; PRT; 609 AA.
ID PRXC_CURIN AC P49053;
DT 01-FEB-1996 (Rel. 33, Created)

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE VANADIUM CHLOROPEROXIDASE (EC 1.11.1.10) (VCPO) (VANADIUM CHLORIDE  
 DE PEROXIDASE).  
 GN CPO.  
 OS Curvularia inaequalis.  
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;  
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Curvularia.  
 [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE; 95262722.  
 RA SIMONS B.H., BARNETT P., VOLLENBROEK E.G.M., DEKKER H.L.,  
 RA MUIJERS A.O., MESSERSCHMIDT A., WEVER R.;  
 RT "Primary structure and characterization of the vanadium  
 RT chloroperoxidase from the fungus Curvularia inaequalis.";  
 RL Eur. J. Biochem. 229:566-574(1995).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
 RX MEDLINE; 96133943.  
 RA MESSERSCHMIDT A., WEVER R.;  
 RT "X-ray structure of a vanadium-containing enzyme: chloroperoxidase  
 RT from the fungus Curvularia inaequalis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:392-396(1996).  
 CC -1- CATALYTIC ACTIVITY: 2 RH + 2 CHLORIDE + H(2)O(2) = 2 RCL + 2  
 CC H(2)O.  
 CC -1- COFACTOR: VANADIUM.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X85369; CAA59686.1; -  
 DR PDB; 1VNC; 08-NOV-96.  
 KW Oxidoreductase; Peroxidase; Vanadium; 3D-structure.  
 FT ACT\_SITE 404 404  
 FT METAL 496 496 VANADIUM.  
 FT CONFLICT 454 454 P -> S (IN AA SEQUENCE).  
 SQ SEQUENCE 609 AA; 67530 MW; 19112E80 CRC32;

Query Match 60.3%; Score 35; DB 1; Length 609;  
 Best Local Similarity 50.0%; Pred. No. 17;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YRLIRRIALRY 12  
 Db 304 YNQIVRIAVTY 315  
 | :|:|:|:|:|  
 RESULT 3  
 RS15\_BORBU STANDARD; PRT; 88 AA.  
 AC Q31744;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 30S RIBOSOMAL PROTEIN S15.  
 GN RPSO OR BB0804.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE; 98065943.  
 RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
 RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
 RA DOUGHERTY B., TOMB J.F., FLEISCHMANN R.D., RICHARDSON D.,

RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
 RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
 RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,  
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
 RA SMITH H.O., VENTER J.C.;  
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
 RT burgdorferi";  
 RL Nature 390:580-586(1997).  
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING  
 CC PROTEINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE001179; AAC67150.1; -  
 DR TIGR; BB0804;  
 DR PROSITE; PS00362; RIBOSOMAL\_S15; 1.  
 DR PFAM; PF00312; Ribosomal\_S15; 1.  
 KW Ribosomal protein; rRNA-binding.  
 SQ SEQUENCE 88 AA; 10217 MW; 054549EE CRC32;

Query Match 58.6%; Score 34; DB 1; Length 88;  
 Best Local Similarity 54.5%; Pred. No. 3.4;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 YRLIRRIALR 11  
 Db 77 YRLISKLGIR 87  
 ||:|:|:|:|  
 RESULT 4  
 RS15\_MYCLE STANDARD; PRT; 89 AA.  
 AC Q32367;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE 30S RIBOSOMAL PROTEIN S15.  
 GN RPSO OR MLCB22.28C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DEVLIN K., CHURCHER C.M., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING  
 CC PROTEINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z98741; CAB11393.1; -  
 DR HSP; P05766; I332.  
 DR PROSITE; PS00362; RIBOSOMAL\_S15; 1.  
 DR PFAM; PF00312; Ribosomal\_S15; 1.  
 KW Ribosomal protein; rRNA-binding.  
 SQ SEQUENCE 89 AA; 10348 MW; 0E24D709 CRC32;

```
Query Match          58.6%; Score 34; DB 1; Length 89;
Best Local Similarity 63.6%; Pred. No. 3.5;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALR 11
DB 78 YRSlierLGLR 88

RESULT 5
RS15_MYCTU STANDARD; PRT; 89 AA.
AC O3327;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S15.
GN RPSO OR RV2785C OR MT0002.50C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE: 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING
CC -!- PROTEINS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL008967; CAA15580.1; -.
DR HSSP: P05766; 1A32.
DR PROSITE: PS00362; RIBOSOMAL_S15; 1.
DR PFAM: PF00312; Ribosomal_S15; 1.
KW Ribosomal protein.
SQ SEQUENCE 89 AA; 10475 MW; 374C7EB8 CRC32;

Query Match          58.6%; Score 34; DB 1; Length 89;
Best Local Similarity 63.6%; Pred. No. 3.5;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRRIALR 11
DB 78 YRSlierLGLR 88

RESULT 6
YHSS_CAEEL STANDARD; PRT; 411 AA.
AC O18304;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 48.2 KD PROTEIN ZK849.5 IN CHROMOSOME I.
DE
```

```
GN ZK849.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA KERSHAW J.;
RC STRAIN-BRISTOL N2;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z82095; CAB05028.1; -.
DR WORMPEP: ZK849.5; CE16750.
DR PFAM: PF01062; Worm_family_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 411 AA; 48236 MW; B5736F70 CRC32;

Query Match          58.6%; Score 34; DB 1; Length 411;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLIRRIALRY 12
DB 128 LLIRRIIIRY 137

RESULT 7
CAO3_CANTR STANDARD; PRT; 502 AA.
AC P11355;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ACYL-COENZYME A OXIDASE POX4-2 (EC 1.3.3.6) (ACYL-COA OXIDASE)
DE (FRAGMENT).
GN POX4-2.
OS Candida tropicalis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Candidaceae; Candida.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87280361.
RA SMALL G.M., LAZAROW P.B.;
RT "Import of the carboxy-terminal portion of acyl-CoA oxidase into
RT peroxisomes of Candida tropicalis."
RL J. Cell Biol. 105:247-250(1987).
CC -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC FROM 8 TO 18).
CC -!- COFACTOR: FAD.
CC -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC SYSTEM.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: PEROXISOMAL.
DR PIR: A28584; OXCKAX.
KW Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW Multigene family.
FT NON_TER 1
SQ SEQUENCE 502 AA; 55528 MW; EACE80C4 CRC32;

Query Match          58.6%; Score 34; DB 1; Length 502;
Best Local Similarity 64.3%; Pred. No. 22;
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;
```

OY 1 YRLIR--RIALRY 12  
 DB 101 YRLARMSTIALRY 114

RESULT 8  
 CAO2\_CANTR STANDARD; PRT; 708 AA.  
 AC P05598;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ACYL-COENZYME A OXIDASE II (EC 1.3.3.6) (ACYL-COA OXIDASE) (PXP-4).  
 GN AOX OR POX-4.  
 OS Candida tropicalis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Candidaceae; Candida.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 20336 / PK233;  
 RX MEDLINE; 87248070.  
 RA MURRAY W.W., RACHUBINSKI R.A.;  
 RT "The primary structure of a peroxisomal fatty acyl-CoA oxidase from  
 RT the yeast Candida tropicalis pK233.";  
 RL Gene 51:119-128(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 20336 / PK233;  
 RX MEDLINE; 86149279.  
 RA OKAZAKI K., TAKECHI T., KAMBARA N., FUKUI S., KUBOTA I., KAMIRYO T.;  
 RT "Two acyl-coenzyme A oxidases in peroxisomes of the yeast Candida  
 RT tropicalis: primary structures deduced from genomic DNA sequence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1232-1236(1986).  
 RN [3]  
 RP SEQUENCE OF 208-709 FROM N.A.  
 RC STRAIN-RII;  
 RX MEDLINE; 87280361.  
 RA SMALL G.M., LAZAROW P.B.;  
 RT "Import of the carboxy-terminal portion of acyl-CoA oxidase into  
 RT peroxisomes of Candida tropicalis.";  
 RL J. Cell Biol. 105:247-250(1987).  
 CC -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +  
 CC H2O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH  
 CC FROM 8 TO 18).  
 CC -!- COFACTOR: FAD.  
 CC -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION  
 CC SYSTEM.  
 CC -!- SUBUNIT: HOMOOCTAMER.  
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M16193; AAA34322.1; -;  
 CC EMBL; M12160; AAA34362.1; -;  
 CC EMBL; Y00623; CAA68660.1; -;  
 CC EMBL; Y00623; CAA68661.1; ALT\_INIT.  
 CC EMBL; Y00623; CAA68662.1; ALT\_INIT.  
 CC PIR; A25123; OXCKX4.  
 CC PIR; A29047; OXCKX.  
 CC Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;  
 CC Multigene family.  
 CC INIT\_MET 0  
 CC Q -> E (IN REF. 3).  
 CC P -> A (IN REF. 2 AND 3).  
 CC N -> K (IN REF. 2).  
 CC FRAMESHIFT ERROR (IN REF. 2 AND 3).

FT CONFLICT 435 435 H -> Y (IN REF. 2).  
 FT CONFLICT 462 462 G -> A (IN REF. 2).  
 FT CONFLICT 495 495 E -> S (IN REF. 2).  
 FT CONFLICT 576 576 ELA -> DLV (IN REF. 2).  
 FT CONFLICT 697 697 Q -> E (IN REF. 2 AND 3).  
 SQ SEQUENCE 708 AA; 79041 MW; D97A4EC8 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 708;  
 Best Local Similarity 64.3%; Pred. No. 33;  
 Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

OY 1 YRLIR--RIALRY 12  
 DB 307 YRLARMSTIALRY 320

RESULT 9  
 CAO4\_CANMA STANDARD; PRT; 708 AA.  
 ID CAO4\_CANMA  
 AC P05335;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 01-MAY-1992 (Rel. 22, Last annotation update)  
 DE ACYL-COENZYME A OXIDASE POX4 (EC 1.3.3.6) (ACYL-COA OXIDASE) (AOX).  
 GN POX4.  
 OS Candida maltosa (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Candidaceae; Candida.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 20184;  
 RX MEDLINE; 88124223.  
 RA HILL D.E., BOULAY R., ROGERS D.;  
 RT "Complete nucleotide sequence of the peroxisomal acyl CoA oxidase  
 RT from the alkane-utilizing yeast Candida maltosa.";  
 RL Nucleic Acids Res. 16:365-366(1988).  
 CC -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +  
 CC H2O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH  
 CC FROM 8 TO 18).  
 CC -!- COFACTOR: FAD.  
 CC -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION  
 CC SYSTEM.  
 CC -!- SUBUNIT: HOMOOCTAMER.  
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X06721; CAA29901.1; -;  
 CC PIR; A29441; OXCKPM.  
 CC Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;  
 CC Multigene family.  
 CC INIT\_MET 0  
 CC Q -> E (IN REF. 2 AND 3).  
 CC P -> A (IN REF. 2 AND 3).  
 CC N -> K (IN REF. 2).  
 CC FRAMESHIFT ERROR (IN REF. 2 AND 3).

Query Match 58.6%; Score 34; DB 1; Length 708;  
 Best Local Similarity 64.3%; Pred. No. 33;  
 Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

OY 1 YRLIR--RIALRY 12  
 DB 307 YRLARMSTIALRY 320

RESULT 10  
 POLG\_DEN26

```

ID POLG_DEN26 STANDARD: PRT: 3391 AA.
AC P29990;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E;
DE NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3);
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)].
OS Dengue virus type 2 (strain 16681).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92188532.
RA BLOK J., MCWILLIAM S.M., BUTLER H.C., GIBBS A.J., WEILLER G.,
RA HERRING B.L., HEMSLEY A.C., AASKOV J.G., YOKSAN S.,
RA BHAMARAPRAVATI N.;
RT "Comparison of a dengue-2 virus and its candidate vaccine derivative:
RT sequence relationships with the flaviviruses and other viruses.";
RL Virology 187:573-590(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84727; AAA73185.1; -
DR PIR; A42451; GNMV16.
DR HSP; P14336; 1SVB.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
DR PFAM; PF01349; Flavi_NS4B; 1.
DR PFAM; PF01350; Flavi_NS4A; 1.
DR PFAM; PF01570; Flavi_propep; 1.
KW Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT CHAIN 1
FT PROPEP 115 205
FT CHAIN 206 280
FT CHAIN 281 775
FT CHAIN 776 1127
FT CHAIN 1128 1345
FT CHAIN 1346 1474
FT CHAIN 1475 2093
FT CHAIN 2094 2243
FT CHAIN 2244 2491
FT CHAIN 2492 3391
FT CHAIN 1668 1675
FT NP_BIND 1759 1762
FT SITE 50 66
FT TRANSMEM 102 118
FT TRANSMEM 268 284
FT TRANSMEM 727 743
FT TRANSMEM 757 773
FT TRANSMEM 1158 1174

```

```

FT TRANSMEM 1272 1288
FT TRANSMEM 1294 1310
FT TRANSMEM 1351 1367
FT TRANSMEM 1373 1389
FT TRANSMEM 1448 1464
FT TRANSMEM 2148 2164
FT TRANSMEM 2174 2190
FT TRANSMEM 2197 2213
FT TRANSMEM 2227 2243
FT TRANSMEM 2352 2368
FT TRANSMEM 2411 2427
FT DISULFID 283 310
FT DISULFID 340 396
FT DISULFID 354 385
FT DISULFID 372 401
FT DISULFID 465 565
FT DISULFID 582 613
FT CARBOHYD 183 183
FT CARBOHYD 347 347
FT CARBOHYD 433 433
FT CARBOHYD 905 905
FT CARBOHYD 982 982
FT CARBOHYD 1134 1134
FT CARBOHYD 2301 2301
FT CARBOHYD 2305 2305
FT CARBOHYD 2457 2457
FT CARBOHYD 2485 2485
FT CARBOHYD 2665 2665
FT CARBOHYD 2704 2704
FT CARBOHYD 2714 2714
FT CARBOHYD 3173 3173
SQ SEQUENCE 3391 AA; 379540 MW; 67D1695A CRC32;

Query Match 58.6%; Score 34; DB 1; Length 3391;
Best Local Similarity 54.5%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLLIRRIALRY 12
DB 2728 RMLINRFTWRY 2738

RESULT 11
POLG_DEN27 STANDARD; PRT: 3391 AA.
AC P29991;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E;
DE NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3);
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)].
OS Dengue virus type 2 (strain 16681-PDK53).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92188532.
RA BLOK J., MCWILLIAM S.M., BUTLER H.C., GIBBS A.J., WEILLER G.,
RA HERRING B.L., HEMSLEY A.C., AASKOV J.G., YOKSAN S.,
RA BHAMARAPRAVATI N.;
RT "Comparison of a dengue-2 virus and its candidate vaccine derivative:
RT sequence relationships with the flaviviruses and other viruses.";
RL Virology 187:573-590(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.

```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M84728; AAA73186.1; -  
DR EMBL; M84727; -; NOT\_ANNOTATED\_CDS.  
DR PIR; B42451; GNWV26.  
DR HSSP; P14336; 1SVB.  
DR PFAM; PF00869; Flavi\_glycoprot; 1.  
DR PFAM; PF00948; Flavi\_NS1; 1.  
DR PFAM; PF00949; Flavi\_helicase; 1.  
DR PFAM; PF00972; Flavi\_NS5; 1.  
DR PFAM; PF01002; Flavi\_NS2B; 1.  
DR PFAM; PF01003; Flavi\_capsid; 1.  
DR PFAM; PF01004; Flavi\_M; 1.  
DR PFAM; PF01005; Flavi\_NS2A; 1.  
DR PFAM; PF01349; Flavi\_NS4B; 1.  
DR PFAM; PF01350; Flavi\_NS4A; 1.  
DR PFAM; PF01570; Flavi\_propep; 1.  
DR Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural  
FT CHAIN 1 114  
FT PROPEP 115 205  
FT CHAIN 206 280  
FT CHAIN 281 775  
FT CHAIN 776 1127  
FT CHAIN 1128 1345  
FT CHAIN 1346 1474  
FT CHAIN 1475 2093  
FT CHAIN 2094 2243  
FT CHAIN 2244 2491  
FT CHAIN 2492 3391  
FT NP\_BIND 1668 1675  
FT SITE 1759 1762  
FT TRANSMEM 50 66  
FT TRANSMEM 102 118  
FT TRANSMEM 268 284  
FT TRANSMEM 727 743  
FT TRANSMEM 757 773  
FT TRANSMEM 1158 1174  
FT TRANSMEM 1272 1288  
FT TRANSMEM 1294 1310  
FT TRANSMEM 1351 1367  
FT TRANSMEM 1373 1389  
FT TRANSMEM 1448 1464  
FT TRANSMEM 2148 2164  
FT TRANSMEM 2174 2190  
FT TRANSMEM 2197 2213  
FT TRANSMEM 2227 2243  
FT TRANSMEM 2352 2368  
FT TRANSMEM 2411 2427  
FT DISULFID 283 310  
FT DISULFID 340 396  
FT DISULFID 354 385  
FT DISULFID 372 401  
FT DISULFID 465 565  
FT DISULFID 582 613  
FT CARBOHYD 183 183  
FT CARBOHYD 347 347  
FT CARBOHYD 433 433  
FT CARBOHYD 905 905  
FT CARBOHYD 982 982  
FT CARBOHYD 1134 1134  
FT CARBOHYD 2301 2301  
FT CARBOHYD 2305 2305  
FT CARBOHYD 2457 2457  
FT CARBOHYD 2485 2485

FT CARBOHYD 2665 2665 POTENTIAL.  
FT CARBOHYD 2704 2704 POTENTIAL.  
FT CARBOHYD 2714 2714 POTENTIAL.  
SQ SEQUENCE 3391 AA; 379878 MW; 70570314 CRC32;  
  
Query Match 58.6%; Score 34; DB 1; Length 3391;  
Best Local Similarity 54.5%; Pred. No. 1.8e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 RLLIRRLRYL 12  
Db 2728 RMLINRFTMY 2738  
  
RESULT 12  
YK50\_YEAST STANDARD; PRT; 352 AA.  
ID YK50\_YEAST AC P36151;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHEORETICAL 39.4 KD PROTEIN IN MET1-SIS2 INTERGENIC REGION.  
GN YKR070W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA POHL T.M., POHL F.M.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z38295; CAAB2149.1; -  
DR PIR; S38147; S38147.  
KW Hypothetical protein.  
SQ SEQUENCE 352 AA; 39406 MW; F5B1BB2A CRC32;  
  
Query Match 56.9%; Score 33; DB 1; Length 352;  
Best Local Similarity 60.0%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 YRLIRRL 10  
Db 223 FRLVRLYL 232  
  
RESULT 13  
YD95\_YEAST STANDARD; PRT; 869 AA.  
ID YD95\_YEAST AC P38966;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHEORETICAL 98.7 KD PROTEIN IN SLU7-BMH2 INTERGENIC REGION.  
GN YDR089W OR D4495 OR YD8554.22.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE OF 1-161 FROM N.A.  
RC STRAIN-S288C / FY1679;  
RA COSTER F., JONNIAUX J.-L., GOFFEAU A.;  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]





Db 133 YRRIIRPIFLRH 144

Search completed: February 8, 2000, 00:59:56  
Job time: 3785 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 13:17:44 ; Search time 209.03 Seconds  
(without alignments)  
3.980 Million cell updates/sec

Title: US-08-653-294-17  
Perfect score: 58  
Sequence: 1 YRLIRIALRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

#### Database :

SPTREMBL12.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	65.5	795	5 Q22063	Q22063 caenorhabdi
2	37	63.8	132	2 Q9X378	Q9X378 bacillus an
3	37	63.8	313	2 P97213	P97213 clostridium
4	37	63.8	1054	4 Q94887	Q94887 homo sapien
5	36	62.1	325	2 Q9XB05	Q9XB05 myxococcus
6	36	62.1	904	4 O15455	O15455 homo sapien
7	35	60.3	309	2 O86347	O86347 mycobacteri
8	35	60.3	490	2 O07074	O07074 porphyromon
9	35	60.3	828	10 O82630	O82630 arabidopsis
10	34	58.6	362	1 O58274	O58274 pyrococcus
11	34	58.6	530	12 Q9IVN7	Q9IVN7 melanoplus
12	34	58.6	3391	12 O11875	O11875 dengue viru
13	34	58.6	3391	12 O09234	O09234 unidentified
14	33	56.9	107	1 O30127	O30127 archaeglob
15	33	56.9	260	12 Q9YW71	Q9YW71 melanoplus
16	33	56.9	550	1 Q9Y821	Q9Y821 aeropyrum p
17	33	56.9	558	5 Q93329	Q93329 caenorhabdi
18	33	56.9	565	2 O53839	O53839 streptomyce
19	33	56.9	593	4 Q9Y483	Q9Y483 homo sapien
20	33	56.9	2304	10 Q42617	Q42617 brassica na

21	32.5	56.0	459	2	032384	032384 synchococc
22	32	55.2	243	2	085853	085853 sphingomona
23	32	55.2	268	2	083438	083438 treponema p
24	32	55.2	274	5	017713	017713 caenorhabdi
25	32	55.2	296	2	047402	047402 escherichia
26	32	55.2	406	2	09X5N8	09X5N8 escherichia
27	32	55.2	408	2	066951	066951 aquifex aeo
28	32	55.2	414	2	069477	069477 mycobacteri
29	32	55.2	452	2	059958	059958 streptococc
30	32	55.2	469	2	083593	083593 treponema p
31	32	55.2	470	2	006317	006317 mycobacteri
32	32	55.2	481	2	09X754	09X754 klebsiella
33	32	55.2	507	10	082580	082580 arachis hyp
34	32	55.2	573	5	097211	097211 leishmania
35	32	55.2	613	3	P79087	P79087 embellisia
36	32	55.2	619	1	09YD18	09YD18 aeropyrum p
37	32	55.2	636	2	006966	006966 bacillus su
38	32	55.2	895	1	09YAU1	09YAU1 aeropyrum p
39	32	55.2	1355	5	016779	016779 caenorhabdi
40	32	55.2	1998	12	056050	056050 rhopalosiph
41	32	55.2	3164	12	09YTU2	09YTU2 cryphonectr
42	32	55.2	3165	12	004350	004350 cryphonectr
43	31	53.4	214	2	0925A0	0925A0 streptomyce
44	31	53.4	248	2	031072	031072 synchococc
45	31	53.4	272	2	P71746	P71746 mycobacteri

#### ALIGNMENTS

RESULT 1  
ID Q22063 PRELIMINARY; PRT; 795 AA.  
AC Q22063; Q93489;  
DT 01-NOV-1996 (TREMREL. 01, Created)  
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)  
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)  
DE T01C3.10 PROTEIN.  
TS T01C3.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WILD A.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z81061; CAB02939.1; -.  
DR EMBL; Z78413; CAB02939.1; JOINED.  
DR EMBL; Z78413; CAB01667.1; -.  
DR EMBL; Z81061; CAB01667.1; JOINED.  
DR HSSP; P19491; 1GR2.  
DR PFAM; PF00060; lig\_chan; 1.  
SQ SEQUENCE 795 AA; 89703 MW; DD722166 CRC32;

Query Match 65.5%; Score 38; DB 5; Length 795;  
Best Local Similarity 58.3%; Pred. No. 31;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLIRIALRY 12  
|||:|:|  
Db 6 YRTSLRLATRY 17  
RESULT 2  
ID Q9X378 PRELIMINARY; PRT; 132 AA.  
AC Q9X378;  
DT 01-NOV-1999 (TREMREL. 12, Created)  
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)  
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)  
DE PX01-117.  
OS Bacillus anthracis.

OG Plasmid virulence plasmid PX01.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STERN;  
 RA OKINAKA R.T., CLOUD K., HAMTON O., HOFFMASTER A., HILL K.K., KEIM P.,  
 RA KOEHLER T., LAMKE G., KUMANO S., MAHILLON J., MANTER D., MARTINEZ Y.,  
 RA RICKE D.O., SVENSSON R., JACKSON P.J.;  
 RT "The sequence and organization of PX01, the large Bacillus anthracis  
 RT plasmid harboring the Anthrax toxin genes.";  
 RL J. Bacteriol. 0:0-0(1999).  
 DR EMBL; AF065404; AAD32421.1; -;  
 KW Plasmid.  
 SQ SEQUENCE 132 AA; 15748 MW; 097E4E16 CRC32;

Query Match 63.8%; Score 37; DB 2; Length 132;  
 Best Local Similarity 63.6%; Pred. No. 8.2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 RLIRRIALRY 12  
 |:|:|:|:|  
 DB 87 RILLRRAALNY 97

RESULT 3  
 ID P97213 PRELIMINARY; PRT; 313 AA.  
 AC P97213;  
 DT 01-MAY-1997 (TREMREL. 03, Created)  
 DT 01-MAY-1997 (TREMREL. 03, Last sequence update)  
 DT 01-MAY-1997 (TREMREL. 03, Last annotation update)  
 DE CDU2, CDU1, TCDD, TCDB, TCDE, TCDA, TCDC, CDD1, CDD2, CDD3, AND CDD4  
 DE GENES.  
 GN CDU2.  
 OS Clostridium difficile.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VPI10463;  
 RA VON EICHEL-STREIBER C.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X92982; CAA63559.1; -;  
 DR EMBL; X92982; CAA63558.1; -;  
 SQ SEQUENCE 313 AA; 33380 MW; 4F20347A CRC32;

Query Match 63.8%; Score 37; DB 2; Length 313;  
 Best Local Similarity 41.7%; Pred. No. 19;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 YRLIRRIALRY 12  
 |:|:|:|:|  
 DB 259 YRVVARRLSIKY 270

RESULT 4  
 ID O94887 PRELIMINARY; PRT; 1054 AA.  
 AC O94887;  
 DT 01-MAY-1999 (TREMREL. 10, Created)  
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)  
 DT 01-NOV-1999 (TREMREL. 12, Last annotation update)  
 DE KIA0793 PROTEIN.  
 GN KIA0793.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;

RX MEDLINE; 99087487.  
 RA NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., MIYAJIMA N., TANAKA A.,  
 RA KOTANI H., NOMURA N., OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XI.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 5:277-286(1998).  
 RN [2]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RX MEDLINE; 99063792.  
 RA SULSTON J.E., WATERSTON R.;  
 RT "Toward a complete human genome sequence.";  
 RL Genome Res. 8:1097-1108(1998).  
 RN [3]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA SUN H., STONEKING T., LANGSTON Y., LAPLANT Y.;  
 RT "The sequence of Homo sapiens BAC clone RG442F18.";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA WATERSTON R.H.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA WATERSTON R.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 1-754 FROM N.A.  
 RA WATERSTON R.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB018336; BAA34513.1; -;  
 DR EMBL; AC005104; AAD1224.1; -;  
 DR HSP; P08567; 1PLS.  
 SQ SEQUENCE 1054 AA; 119888 MW; 3BA89171 CRC32;

Query Match 63.8%; Score 37; DB 4; Length 1054;  
 Best Local Similarity 58.3%; Pred. No. 63;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 YRLIRRIALRY 12  
 |:|:|:|:|  
 DB 688 YRLIRRLCGHY 699

RESULT 5  
 QXAB05 PRELIMINARY; PRT; 325 AA.  
 ID Q9XB05  
 AC Q9XB05;  
 DT 01-NOV-1999 (TREMREL. 12, Created)  
 DT 01-NOV-1999 (TREMREL. 12, Last sequence update)  
 DT 01-NOV-1999 (TREMREL. 12, Last annotation update)  
 DE MEMBRANE ASSOCIATED PROTEIN.  
 GN TAD.  
 OS Myxococcus xanthus.  
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;  
 OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ER-15;  
 RA PAITAN Y., ORR E., RON E.Z., ROSENBERG E.;  
 RT "Genetic and functional analysis of genes required for the post-  
 RT modification of the polyketide antibiotic TA of Myxococcus xanthus.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ132503; CAB46503.1; -;  
 SQ SEQUENCE 325 AA; 35985 MW; 4CC64E85 CRC32;

Query Match 62.1%; Score 36; DB 2; Length 325;  
 Best Local Similarity 58.3%; Pred. No. 31;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 YRLIRRIALRY 12

Db 202 YRLVDRPLRY 213  
 ||| : | |||

## RESULT 6

ID O15455 PRELIMINARY; PRT: 904 AA.  
 AC O15455;  
 DT 01-JAN-1998 (TREMELrel. 05, Created)  
 DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE TOLL-LIKE RECEPTOR 3.  
 GN TLR3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA ROCK F.L., HARDIMAN G., TIMANS J.C., KASTELEIN R.A., BAZAN J.F.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U88679; AAC34134.1; -  
 DR PFAM: PF00560; LRR; 15.  
 DR PFAM: PF01463; LRRCT; 1.  
 DR PFAM: PF01582; TIR; 1.  
 DR PRINTS: PRO00019; LEURICRPT.  
 SQ SEQUENCE 904 AA; 103828 MW; F857CE1C CRC32;

Query Match 62.1%; Score 36; DB 4; Length 904;

Best Local Similarity 60.0%; Pred. NO. 85;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLLIRIALR 11  
 |||:|||||  
 Db 484 RMLRRVALK 493

## RESULT 7

ID O86347 PRELIMINARY; PRT: 309 AA.  
 AC O86347;  
 DT 01-NOV-1998 (TREMELrel. 08, Created)  
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE HYPOTHETICAL 33.5 KD PROTEIN.  
 GN RV276C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA COLE S.T., BROSC R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,  
 RA GORDON S.V., EIGMEIER K., GAS S., BARRY III C.E., TEKAIA F.,  
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,  
 RA DAVIES R., DEVLIN K., FLETCHER T., GENTILES S., HAMLIN N., HOLROYD S.,  
 RA HORNSBY T., JAGELS K., KROCH A., MCLEAN J., MOULE S., MURPHY L.,  
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,  
 RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SOARES R., SULSTON J.E.,  
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-H37RV;  
 RA PARKHILL J.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL008967; CAAL5591.1; -  
 DR HSSP: P33164; 2PIA.  
 DR PROSITE: PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PFAM: PF00111; fer2; 1.

DR PFAM: PF00175; oxidored\_fad; 1.  
 KW Hypothetical protein; Iron-sulfur.  
 SQ SEQUENCE 309 AA; 33517 MW; B152B590 CRC32;

Query Match 60.3%; Score 35; DB 2; Length 309;  
 Best Local Similarity 77.8%; Pred. NO. 46;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YRLIRRIA 9  
 ||:|||||  
 Db 65 YRIAIRRIA 73

## RESULT 8

ID O07074 PRELIMINARY; PRT: 490 AA.  
 AC O07074;  
 DT 01-JUL-1997 (TREMELrel. 04, Created)  
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)  
 DE FIMBRILIN.  
 GN FIMA.  
 OS Porphyromonas gingivalis (Bacteroides gingivalis).  
 OC Bacteria; Cytophagales; Bacteroidaceae; Porphyromonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA HONGO H.;  
 RT "Nucleotide sequence upstream from the fimbrilin gene (fima) of the  
 oral anaerobe Porphyromonas gingivalis."  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB004560; BAA33066.1; -  
 SQ SEQUENCE 490 AA; 55272 MW; 385AD4FF CRC32;

Query Match 60.3%; Score 35; DB 2; Length 490;  
 Best Local Similarity 70.0%; Pred. NO. 72;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 YRLIRRIAL 10  
 ||:|||||  
 Db 92 YRIIDRIAL 101

## RESULT 9

ID O82630 PRELIMINARY; PRT: 828 AA.  
 AC O82630;  
 DT 01-NOV-1998 (TREMELrel. 08, Created)  
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
 DE STELAR K+ OUTWARD RECTIFYING CHANNEL.  
 GN SKOR.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV. LANDSBERG ERECTA;  
 RX MEDLINE: 98412660.  
 RA GAYMARD F., PILOT G., LACOMBE B., BOUCHEZ D., BRUNEAU D.,  
 RA BOUCHEZ J., MICHAUX-FERRIERE N., THIBAUD J.B., SENTENAC H.;  
 RT "Identification and disruption of a plant shaker-like outward channel  
 involved in K+ release into the xylem sap."  
 RL Cell 94:647-655(1998).  
 DR EMBL: AJ223358; CAAL1281.1; -  
 DR HSSP: Q00421; 1AWC.  
 DR PFAM: PF00023; ank; 3.  
 DR PFAM: PF00914; CNG\_membrane; 1.  
 DR PFAM: PF00027; CNMP\_binding; 1.

SQ SEQUENCE 828 AA; 93898 MW; AA41E3DD CRC32;

Query Match 60.3%; Score 35; DB 10; Length 828;

Best Local Similarity 57.1%; Pred. No. 1.2e+02;

Matches 8; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

OY 1 YRLIR--IALRY 12

|||||

Db 143 YRMYKRSIALRY 156

RESULT 10

O58274

ID O58274 PRELIMINARY; PRT; 362 AA.

AC O58274;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)

DE 362AA LONG HYPOTHETICAL PROTEIN.

GN PH0539.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3.

RX MEDLINE; 98344137.

RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,

RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,

RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,

RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,

RA AKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,

RA KIKUCHI H.,

RT "Complete sequence and gene organization of the genome of a hyper-

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RL DNA Res. 5:55-76(1998).

DR EMBL; AP000002; BAA29628.1; -.

SQ SEQUENCE 362 AA; 41831 MW; EB8D18CF CRC32;

Query Match 58.6%; Score 34; DB 1; Length 362;

Best Local Similarity 50.0%; Pred. No. 83;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 YRLIRIRIALRY 12

|||||

Db 296 YRFLKIALGY 307

RESULT 11

O9YVN7

ID O9YVN7 PRELIMINARY; PRT; 530 AA.

AC O9YVN7;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)

DE ORF MSV205 TRYPTOPHAN REPEAT GENE FAMILY PROTEIN.

GN MSV205.

OS Melanoplus sanguinipes entomopoxvirus.

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TUCSON;

RX MEDLINE; 99102612.

RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;

RT "The genome of Melanoplus sanguinipes entomopoxvirus.";

RL J. Virol. 73:533-552(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=TUCSON;

RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF063866; AAC97756.1; -.

SQ SEQUENCE 530 AA; 65062 MW; D29ED6FE CRC32;

Query Match 58.6%; Score 34; DB 12; Length 530;

Best Local Similarity 50.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 YRLIRIRIALRY 12

|||||

Db 139 HRLYLREIPRY 150

RESULT 12

O11875

ID O11875 PRELIMINARY; PRT; 3391 AA.

AC O11875;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE POLYPROTEIN.

OS Dengue virus type 2.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PDK-53.

RX MEDLINE; 97288308.

RA KINNEY R.M., BUTRAPET S., CHANG G.-J., TSUCHIYA K.R., ROEHRIG J.T.,

RA BHAKRAPRATVI N., GUBLER D.J.;

RT "Construction of infectious cDNA clones for dengue 2 virus: strain

RT 16681 and its attenuated vaccine derivative, strain PDK-53.";

RL Virology 230:300-308(1997).

DR EMBL; U87412; AAB58783.1; -.

DR HSP; P4336; ISVB.

DR PFAM; PF00869; Flavi\_glycoprot; 1.

DR PFAM; PF00949; Flavi\_helicase; 1.

DR PFAM; PF01004; Flavi\_M; 1.

DR PFAM; PF00948; Flavi\_NS1; 1.

DR PFAM; PF01005; Flavi\_NS2A; 1.

DR PFAM; PF01002; Flavi\_NS2B; 1.

DR PFAM; PF01350; Flavi\_NS4A; 1.

DR PFAM; PF01349; Flavi\_NS4B; 1.

DR PFAM; PF00972; Flavi\_NS5; 1.

DR PFAM; PF01570; Flavi\_proteg; 1.

DR PFAM; PF01003; Flavi\_capsid; 1.

KW Polyprotein.

SQ SEQUENCE 3391 AA; 379861 MW; 5F56B205 CRC32;

Query Match 58.6%; Score 34; DB 12; Length 3391;

Best Local Similarity 54.5%; Pred. No. 7.5e+02;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 RLLIRIRIALRY 12

|||||

Db 2728 RMLINRFTMY 2738

RESULT 13

O09234

ID O09234 PRELIMINARY; PRT; 3391 AA.

AC O09234;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE POLYPROTEIN.

OS unidentified.

OC unclassified.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-16681;

RX MEDLINE; 97288308.

RA KINNEY R.M., BUTRAPET S., CHANG G.J., TSUCHIYA K.R., ROEHRIG J.T.,  
 RA BHARAPRAVATI N., GUBLER D.J.;  
 RT "Construction of infectious cDNA clones for dengue 2 virus: strain  
 RT 16681 and its attenuated vaccine derivative, strain PDK-53";  
 RL Virology 230:300-308(1997).  
 DR EMBL; U87411; AAB58782.1; -;  
 DR HSSP; P14336; ISVB.  
 DR PFAM; PF00869; Flavi\_glycoprot; 1.  
 DR PFAM; PF00949; Flavi\_helicase; 1.  
 DR PFAM; PF01004; Flavi\_M; 1.  
 DR PFAM; PF00948; Flavi\_NSI; 1.  
 DR PFAM; PF01005; Flavi\_NS2A; 1.  
 DR PFAM; PF01002; Flavi\_NS2B; 1.  
 DR PFAM; PF01350; Flavi\_NS4A; 1.  
 DR PFAM; PF01349; Flavi\_NS4B; 1.  
 DR PFAM; PF00972; Flavi\_NS5; 1.  
 DR PFAM; PF01570; Flavi\_propep; 1.  
 DR PFAM; PF01003; Flavi\_capsid; 1.  
 KW Polyprotein.  
 SQ SEQUENCE 3391 AA; 379800 MW; A839A51C CRC32;

Query Match 58.6%; Score 34; DB 12; Length 3391;  
 Best Local Similarity 54.5%; Pred. No. 7.5e+02;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 RLIRRIALRY 12  
 | : | | : | |  
 DB 2738 RMLNRFTRY 2738

RESULT 14  
 O30127 PRELIMINARY; PRT; 107 AA.  
 AC O30127;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN.  
 GN AF0109.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE; 98049343.  
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,  
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,  
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,  
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,  
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,  
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,  
 RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,  
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,  
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,  
 RA VENIER J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing Archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 DR EMBL; AE001099; AAB91122.1; -;  
 DR TIGR; AF0109; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 107 AA; 13307 MW; F23652A7 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 107;  
 Best Local Similarity 58.3%; Pred. No. 39;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YRLIRRIALRY 12  
 | | | : | | |

DB 38 YVELARKIAMRY 49  
 RESULT 15  
 Q9YW71 PRELIMINARY; PRT; 260 AA.  
 ID Q9YW71  
 AC Q9YW71;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE ORF MSV021 MTG MOTIF GENE FAMILY PROTEIN.  
 GN MSV021.  
 OS Melanoplus sanguinipes entomopoxvirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
 OC Entomopoxvirus B.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TUCSON;  
 RX MEDLINE; 99102612.  
 RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;  
 RT "The genome of Melanoplus sanguinipes entomopoxvirus.";  
 RL J. Virol. 73:533-552(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TUCSON;  
 RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF063866; AAC97850.1; -;  
 SQ SEQUENCE 260 AA; 31180 MW; AB880E14 CRC32;

Query Match 56.9%; Score 33; DB 12; Length 260;  
 Best Local Similarity 58.3%; Pred. No. 93;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 YRLIRRIALRY 12  
 | | | | : | | |  
 DB 234 YLLIEKILNY 245  
 Search completed: February 8, 2000, 13:17:45  
 Job time: 32494 sec

**THIS PAGE BLANK (USPTO)**



[illegible]





```

Mycobacterium.
1 (bases 1 to 38110)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov
12:396(6707):190]]
2 (bases 1 to 38110)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams: Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1261921.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1. 38110
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
<1. 5460
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
/clones="V190"
163..1599
/gene="glnA1"
163..1599
/gene="glnA"
/gene="glnA"
/feature="Rv2220, (MTCY190_31-MTCY427.01), len: 478, glnA1,
Probable glutamine synthetase, glnA, similar to many eg
SW:GLNA_STRC P15106 (71.4% identity in 475 aa overlap);
contains PS00180 Glutamine synthetase signature 1, PS00181
Glutamine synthetase putative ATP-binding region
signature, PS00182 Glutamine synthetase class-I
adenylation site"
/codon_start=1
/transl_table=11
/product="glnA"
/protein_id="CAA94665.1"
/db_xref="GI:3261568"
/db_xref="SWISS-PROT:Q10377"
/translation="MEKTEPDDVFKLAKDEKVEVDVRCDFLPGIMQHETIPASAFDK
SVDFDGLAFGSSIRGFQSIHESDMLLPDPETARIDPFRAAKTLNINFFVHDFPFI
SYRDPNRIKAEENYLIISTADIATYFAGAEAFYIFDVSFDSFRANGSYFYEVDIISG
WNTGAATEADGSPNRCYKVRHKGFFVPVANDQYVDLRDKMLTNLINSFILERGHH
EYVSGSQAEINQFNLSLHAADDMLQYIKYIKNTAWQNGKTVTFMPKPLFGDNGSGMH
CHQSLWKDGAPLMYDETGYAGLSDTARHYVIGLLHAPSLAFTPTNPTVNSYKRLPGY
EAPINLYSQNRSAVRIPITGSPNPKAKLEFRSPDSSGPNYLAFSAMLAGLGIK
NWFEPQPVKDLIYELPPEEAASIFQTPQLQSDVIDRLEADHEYLTEGGVFTNDLIET
WISFKREIEPVIINRPHYFALYDYV"
382..>38110
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clones="Y427"
955..1002
/gene="glnA1"
/feature="P500181 Glutamine synthetase putative ATP-binding
region signature"
1342..1380
/gene="glnA1"
/feature="P500182 Glutamine synthetase class-I adenylation
site"
complement(1917..4901)
/gene="glnE"
complement(1917..4901)
/gene="glnE"
/feature="Rv2221c, (MTCY190_32c-MTCY427.02c), len: 994.
glnE, Probable glnE, similar to SW:GLNE_ECOLI P30870
glutamate-ammonia-ligase adenylyltransferase (24.4%
identity in 721 aa overlap) and SW:GLNE_HAEIN P44419 (28.1%
identity in 199 aa overlap); initiation codon uncertain
unless at 39346 or 39247"
/codon_start=1
/transl_table=11
/product="glnE"
/protein_id="CAA94664.1"
/db_xref="GI:1261923"
/db_xref="SWISS-PROT:Q10379"
/translation="MVYTKATQPKLPVSGRLGLVDPGALRQAQLGWDHEDQAHV
DLWLSRAPDADAALRALIRLSENPDTGDELNAALRLRSLRGLFSVLGSSLAIG
DHLVAHPQSKLLRGKVTLPDQDSFECVEESEGPGSLVHLRRTQYRDYVLM
ADLDAATVEDEPVLPTVVAARLADAADAALAAALRVAAESVCGEHPRLAVIANG
KGAERLNVSDVDVIFVAERSDPRNARVASAEFEVDAALRPEGRNGFLV
RLESHIAYQWAKTWERQALLKARPVGDALGERYLTALMPWVRACERDEFLV
VAMKRRQLPADVRGRELKGGSDRDFAVQLQLVHARDESURASTVDAL
AALGGGYIGREDAANMTASYEFLLEHLRQLKRLKTHLLPDEDEAEAWLRAAA
HIRPDNDADAGLREELKQNVRSKLTFLQPLLESIGTLEIAHGTLEAAG
LSEALATESWTLATLRDKPAVAKRLMVLGTSAYVVDLLMRAPVTOQYEDCGAPGKL
LETEPAVARALIASRYPDPERATAGARTIRRELARIGSADILLGLEVTEVCRL
TSWVAVLQALDVIRASLPDDDRAPAIAVIGRGLGAEGLGSDADVMFCEPA
TGVDARAVKSTISAERVALLGTPSDPPELDANLRPEGRNGPLVRLTGYAAY
EQWAPWEIQALLRAHVAAGDELQORFLRMVDKTRYPPDGVSADSVREIRKARIE
SERLPRGADPNTHTKLGRGLADIENTVQLLOLQHAHQVPALHNTSTLOSVDVIAAD
LVPAADVELLQAWLTATARNALVLVRGKPTDQLPGGRQLNVAVAAGWRNDDGGE
FLDNYLVTRRAKAVRVKFGS"
complement(4950..5290)
/gene="glnA2"
complement(4950..5290)
/gene="glnA2"
/feature="Rv2222c, (MTCY427.03c), len: 446, glnA2, probable
gln2, (overlaps MTCY190_33c partial sequence) probable
classII glutamine synthetase, similar to MTCY190_31 (and
partialorf MTCY427.01) (34.3% identity in 300 aa overlap)"
/codon_start=1
/transl_table=11
/product="glnA2"
/protein_id="CAA94645.1"
/db_xref="GI:1261924"
/db_xref="SWISS-PROT:Q10378"
/translation="MDRQKEFVRLTEERDIFRVLWFTDVLGFLKSAVAIAPELEGA
FEEGICDFGDSIEGFARVSESDTVAHPDSTFQVLPWATSSGHHHSARFCDITPDD
SPSWADPRVLRLRQLTKAGELGFCYVHPHEIEFFLLKPGPDGCVVPVDNAGYDQA
VDSALNFRRHAIADALEFMGISVEFSHHEGAPGOEIDLRFADALSMADNVMVTRFVI
KEVALLEGARASFEMKPFQCHPGSAMHMFLEFGDVNAFHSADLPLOLVSKGFFIA
GILEHACEISAVTNQWNSYKRLVOGGENAPTASGAANRSLVRVPMTPKHTSSRR
VEVRSPDSACNPYLTFVALLAAGLRGKVEKGYVLGPOAEDNVWDLTPERRAMGYREL

```

```
SSLDALRAMEASELVAEALGSEHVFDFFLRNKRRTWANYRSHVTPYELRYLSL"
6349..6366
/Note="inverted repeat between 3' end of MTCY427.04c and
5' end of MTCY427.03c"
complement(6385..7947)
/gene="Rv2223c"
complement(6385..7947)
/gene="Rv2223c"
/Note="Rv2223c. (MTCY427.04c), len: 520. Probable
exported protease, has signal sequence very similar to
three proteases/peptidases from Streptomyces spp.: L42758,
L42759, L27456. FASTA score: L42758|STMSLPD STMSLPD NID:
9940302 - Streptomyces (539 aa) opt: 1032 E(): 0. (37.5%
identity in 533aa overlap). Also similar to hypothetical
proteins SW:YZZE_ECOLI P34211 (25.4% identity in 406 aa
overlap) and PR:B36944 in omp3 3' region (27.5% identity
in 218 aa overlap). Highly similar to to MTCY427.05c
(49.3% identity in 507 aa overlap); contains PS00120
Lipases, serine active site"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2223c"
/protein_id="CAA94646.1"
/db_xref="GI:1261925"
/db_xref="SWISS-PROT:Q10508"
/translation="MAAMWRRPLSSALLSFGLLGLPLAAPPAGATEEPGAGQTP
GAPVAPQSNCSREFIADTSEITARCATVSPVDYDQPGTQAKLAVIRPATGO
REGALLVNPQPGASAVDMVAMAPAIADTDLRFDLVGDPGRGVSTPALRCRTD
AEFDAYRDMADYSFAGVTHVEQYRQLAQCDVDMGFSFLANIGTASVAKMDMVR
QALGDQNLNIGYSYGTGLTAYLERFTHVAMVLDGAIDPANSPIEESIQMAGFQ
TAFNDYAADCARSPACPLGTSAQWNRHYHALVPLVQPKGTSDPRLSYADATGT
INALYSPQRKYLTSGLLQGRSDAGLLVLDYDGRDADGHYSNQDADFNAVRCV
DAPTADPAAWAADRIQVAPFSLSYQGTGSAPRDLCALPWPVATSTPHPAAPAGA
GKVYVSTHDPATPYQSGVDLARQLAPLIIFDGTQHTAFDGNQCVDSAVMHVFLD
GLTPTSURCAP"
complement(7243..7272)
/gene="Rv2223c"
/Note="PS00120 Lipases, serine active site"
complement(7952..7956)
/Note="GGTGA, possible ribosome binding site upstream
Rv2223c"
complement(8009..9571)
/gene="Rv2224c"
complement(8009..9571)
/gene="Rv2224c"

alignment_scores:
  Quality: 42.00 Length: 12
  Ratio: 4.200 Gaps: 0
  Percent Similarity: 83.333 Percent Identity: 75.000

alignment_block:
  US-08-653-294-17 x MTCY427/rev ..
  Align seg 1/1 to reverse of: MTCY427 from: 1 to: 38110
  1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
  |||||:|||||:|||||:|||||:|||||:|||||:
  4941 TACCGGTGTATCCGACCGCCGCGCTTGGCGTAC 4906

seq_name: gb_bai:SS100KBFR
seq_documentation_block:
LOCUS SS100KBFR 100389 bp DNA BCT 29-APR-1997
DEFINITION S.solfataricus 100 kbp DNA fragment.
ACCESSION Y08256
VERSION Y08256.1 GI:1707679
KEYWORDS acylamino-acid-releasing enzyme; alpha-amylase; amidase;
aminopeptidase; daunorubicin resistance protein; glycerol kinase;
glycogen operon protein G1g; H1t protein; IS element; L-lactate
permease; oxidoreductase; pyruvate-ferredoxin oxidoreductase;
riboflavin biosynthesis protein; sugar transporter.
SOURCE Sulfolobus solfataricus.
```

```
ORGANISM Sulfolobus solfataricus
REFERENCE 1 (bases 1 to 100389)
AUTHORS Sensen,C.W., Klenk,H.P., Singh,R.K., Allard,G., Chan,C.C.,
Liu,Q.Y., Penny,S.L., Young,F., Schenk,W.E., Gaasterland,T.,
Doolittle,W.F., Ragan,M.A. and Charlebois,R.L.
Organizational characteristics and information content of an
archaeal genome: 156 kb of sequence from Sulfolobus solfataricus P2
Mol. Microbiol. 22 (1), 175-191 (1996)
JOURNAL 97055432
MEDLINE 97055432
AUTHORS Sensen,C.W.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-1996) C.W. Sensen, Institute for Marine
Biosciences, National Research Council of Canada, 1411 Oxford
Street, Halifax, N.S., Canada, B3H 3Z1, CANADA
COMMENT Related sequence: Y08257.
FEATURES
    source
        1..100389
            /organism="Sulfolobus solfataricus"
            /strain="P2"
            /db_xref="taxon:2287"
            complement(889..1632)
            /gene="orf c06001"
            complement(889..1632)
            /gene="orf c06001"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA69509.1"
            /db_xref="GI:1707680"
            /db_xref="SPTREMBL:P95849"
            /translation="MMIFROIISKGGCATVFGTQAGELFVVPKYMVDEIIEILA
            RLDGMKTAYIIDTHADHLSGAKKLASLNANVYVHELSQLKVKYRDKVSELDYVAGNVSYSWS
            GNVKVLHTPGHTPDSISVLYDRRDESNWPAITGTLFVGGGLGRDIDIGENA
            BENYISLAKELPDIYEIYIPHTAGSVCGIGISGPSSTIGFEKFNLSFRNKND
            EFNRIREVMPKPEFDYIRKNLEGGVI"
            complement(1664..2335)
            /gene="orf c06002"
            complement(1664..2335)
            /gene="orf c06002"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA69510.1"
            /db_xref="GI:1707681"
            /db_xref="SPTREMBL:P95850"
            /translation="MNSLEINLRKLGFSVYEAKVYLTLLNCLNCKMKELSENAEIPY
            QKVEYVKSLEDKGLIVRYIEGRPKVKLIDPSISLKYVRDKVSELDYVAGNVSYSWS
            EKGGEADRSIHKGITVMRMRELVDKSNKIKVYWDILPEWLKILDKKYGKNTVI
            TSSNNLSLNAEVYKYNKIKSFIIIFDQSVLTFNDQDEIVVDSCRGCVLQAEHFDDL
            TYKSE"
            complement(2360..2665)
            /gene="orf c06003"
            complement(2360..2665)
            /gene="orf c06003"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA69511.1"
            /db_xref="GI:1707682"
            /db_xref="SPTREMBL:P95851"
            /translation="MMTVYMLIERNKGFTEINATNPITSTLFLPKSSSESFPAMKLNAP
            FTNPKDTKGNRLNIIEIFCSNTGIIILRPYRNPVKNTNIVNKNCLFIISTKI"
            2376..3476
            /gene="orf c06004"
            2376..3476
            /gene="orf c06004"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA69512.1"
            /db_xref="GI:1707683"
            /db_xref="SPTREMBL:P95852"
            /translation="MMNRQFILTILVFTTGLYGLTLRIIPVEKQINISIMLSLLL
            PLVSFGYVKGFANFIAGKLSDDLGRKRVLIGWALISVPLFSLINITYVTIISILL
            AINQALTWTITVTSQIDISGKLRAFATGINEMSGYLGVSFGSLFASFLKLSGIFIG
```

IICIALISFNVIETKTIPNATLSKKNHINIFGTRKISAGLKEKFDVSAFFIL  
IPTFLLOHYFLIGITVSSYTWLSOPFGYADTYNKRRLIIVIGFLMFVGF  
IKYSELPITLSIIGGIMGYPNLIAFVNDKINESYRGALGYRLYRDSGYVAGL  
LLPLLYSFYGYEYLLVIGILOVALLLVRS"  
complement(3665..4024)  
/gene="orf c06005"  
complement(3665..4024)  
/gene="orf c06005"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA69513.1"  
/db\_xref="GI:1707684"  
/db\_xref="SPTREMBL:P95853"  
/translation="MRRVSKNCNGGKRNKRVLLDALGKNNYKNSVKVYIESIPAIL  
IVSLMKGYSIIINFALTLLSLSPVIAIPAVLVGILISNKRINGKYAYTKRLLAYF  
LTVAMIVAGGLVGLHL"  
complement(3975..5039)  
/gene="orf c06006"  
complement(3975..5039)  
/gene="orf c06006"  
/note="IS element ISC1217"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA69514.1"  
/db\_xref="GI:1707685"  
/db\_xref="SPTREMBL:P95854"  
/translation="MTKELREYVYKALEKAVNEVILSMTGRKDVAKRLLLIGAVVGR  
NATEIAQEMDYETVNLNNDKAAQKALIEVVKLVGHPVLLIVDTHDKLYARAM  
PVSRNGAECFYCRAHKRFPAIQLLVIGVDLVNNOIYVHIIAYIPRKVEEELKRRG  
EVPFKKDALLEFLSLGKLVKARFVDSYNSRTLGNTVGLKSSARVVEGR  
SVYSEFPQGVLYGTIKLVLDIDYKGYRRYFSTDNDTAEIDIITWENRWD  
IEVLRELKALGEGGSLFTWRNSGFVALKALLVQVYFKYSTGLMLGAKRLARI  
KSIYREAGGINKLFRKKRP"  
complement(5112..5990)  
/gene="orf c06007"  
complement(5112..5990)  
/gene="orf c06007"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA69490.1"  
/db\_xref="GI:1707686"  
/db\_xref="SPTREMBL:P95854"  
/translation="MSIRDSAKLFGPAWIALADADAASILGGLSTGEYGYKLWVF  
MLSLPIFIQESGRGAGIGELIREHSTRISILSIFPILVDFEFTYLSYA  
GIAIGCVLIGVPLFVLVFLHVLVKNYEITEKVLWISVILLALIMVGP  
LSFTGQDVFEVSTKFLFIATVNVGAVTPPCMVLVQSSATAIKYSKIGIDNSKTS  
WIGETLLGAIITELIIVLSEMIGTSIGNIDPIDPAQLLNLGSHVHIFITILISAGP  
LTLIVLSLSSANGWGLE"  
6315..6881  
/gene="orf c06008"  
6315..6881  
/gene="orf c06008"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA69491.1"  
/db\_xref="GI:1707687"  
/db\_xref="SPTREMBL:P95855"  
/translation="MQEKQPEKVVGVVEILEKSLDLIKLVLDKLVKATAAEFTYVY  
TILRWHTMEGEGKLEIAEDARLEHFLMFORIYELGGGLPDRLOLAISACS  
DAYLPENWKPKEILKVLAEQCAIRTWEGCDMTYGDPTDYDLAQRIQLBEIEHE  
AWFLELYGRPSGHFRSPGNAPYSKK"  
6934..7260  
/gene="orf c06009"  
6934..7260  
/gene="orf c06009"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA69492.1"  
/db\_xref="GI:1707688"  
/db\_xref="SPTREMBL:P95856"  
/translation="MLVRVSLDLEKKPKPSINGIEVIVRVGDKVFAIAYCPH  
KGNRLYGEVEGYKIRCDLHGYSLENGELIINPNYKTSWTFSPNRLRYKVEIKG  
DIFIQI"

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

alignment\_scores:

Quality:	42.00	Length:	11
Ratio:	4.200	Gaps:	0
Percent Similarity:	90.909	Percent Identity:	81.818

alignment\_block:

US-08-653-294-17 x SS100KBFR/rev ..

Align seg 1/1 to reverse of: SS100KBFR from: 1 to: 100389

1 TyrArgLeuLeuileAargileAalaLeuArg 11  
|||||  
39542 TACITACTCCTAATTCGCGGTTTGCCTTAAGA 39510

seq\_name: gb\_htg3:AC009602

seq\_documentation\_block:

LOCUS	AC009602	40000 bp	DNA	HTG	02-OCT-1999
DEFINITION	Leishmania major chromosome 35 clone L2259 strain Friedlin, ***				
SEQUENCING IN PROGRESS	***, 6 unordered pieces.				
ACCESSION	AC009602				
VERSION	AC009602.2	GI:5006095			
KEYWORDS	HTG; HTGS_PHASE1.				
SOURCE	Leishmania major.				
ORGANISM	Leishmania major				
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;				
REFERENCE	1 (bases 1 to 40000)				

AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrook,M.,  
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.

TITLE Direct Submission

JOURNAL Submitted (28-AUG-1999) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

COMMENT On Oct 1, 1999 this sequence version replaced gi:5801666.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2595: contig of 2595 bp in length  
\* 2596 2875: gap of unknown length  
\* 2876 4361: contig of 1486 bp in length  
\* 4362 4641: gap of unknown length  
\* 4642 8533: contig of 3892 bp in length  
\* 8534 8813: gap of unknown length  
\* 8814 14491: contig of 5678 bp in length  
\* 14492 14771: gap of unknown length  
\* 14772 22462: contig of 7691 bp in length  
\* 22463 22741: gap of unknown length  
\* 22742 40000: contig of 17259 bp in length.

FEATURES  
source  
1. 40000  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="35"  
/clone="L2259"

BASE COUNT 9346 a 10082 c 10936 g 8195 t 1441 others

ORIGIN  
alignment\_scores  
Quality: 41.00 Length: 12  
Ratio: 3.727 Gaps: 0  
Percent Similarity: 91.667 Percent Identity: 66.667

alignment\_block:  
US-08-653-294-17 x AC009602/rev ..  
Align seg 1/1 to reverse of: AC009602 from: 1 to: 40000  
1 TyrArgLeuLeuLeuArgArgIleAlaLeuArgTyr 12  
|||||  
28900 TATCGTCTGCTCATCTGCTGCTGTTGTTGTTCTTTT 28865

seq\_name: gb\_hgt1.CNS01DTM

seq\_documentation\_block:  
LOCUS CNS01DTM 191442 bp DNA HTG 04-NOV-1999  
DEFINITION Homo sapiens chromosome 14 clone R-80A15, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in ordered pieces.

ACCESSION AL132800  
VERSION AL132800.1 GI:6272127  
KEYWORDS HTG; HTGS-PHASE2.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 191442)

REFERENCE Genoscope.

AUTHORS Direct Submission

TITLE Submitted (04-NOV-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continue. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector,  
phage, etc.

Contig order : 13 10 12 11 14 15 9, 1000 N's separate segments  
Contig 13 : length 15976 bp  
Contig 10 : length 5980 bp  
Contig 12 : length 9512 bp  
Contig 11 : length 8124 bp  
Contig 14 : length 67723 bp  
Contig 15 : length 81177 bp  
Contig 9 : length 4117 bp.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1. 191442  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone\_lib="RPCI-11"  
/clone="R-80A15"  
67422 67594  
/standard\_name="M69137"  
/note="matching EMBL:M69137; Identified using the e-PCR  
software (G. Schuler)"  
132481 132607  
/standard\_name="AA194835"  
/note="matching EMBL:AA194835; Identified using the e-PCR  
software (G. Schuler)"  
133712 133887  
/standard\_name="241390"  
/note="matching EMBL:241390; Identified using the e-PCR  
software (G. Schuler)"  
133741 133892  
/standard\_name="R39150"  
/note="matching EMBL:R39150; Identified using the e-PCR  
software (G. Schuler)"  
133768 133909  
/standard\_name="T31019"  
/note="matching EMBL:T31019; Identified using the e-PCR  
software (G. Schuler)"  
146749 146899  
/standard\_name="R37382"  
/note="matching EMBL:R37382; Identified using the e-PCR  
software (G. Schuler)"  
146875 146986  
/standard\_name="R52665"  
/note="matching EMBL:R52665; Identified using the e-PCR  
software (G. Schuler)"  
154585 154748  
/standard\_name="T15974"  
/note="matching EMBL:T15974; Identified using the e-PCR  
software (G. Schuler)"  
BASE COUNT 53877 a 41528 c 39774 g 50165 t 6098 others  
ORIGIN

alignment\_scores:  
Quality: 41.00 Length: 12  
Ratio: 3.727 Gaps: 0  
Percent Similarity: 91.667 Percent Identity: 58.333

alignment\_block:

US-08-653-294-17 x CNS01DTM ..

Align seg 1/1 to: CNS01DTM from: 1 to: 191442

1 TyrArgLeuLeuLeuArgArgIleAlaLeuArgTyr 12  
|||||  
1830 TACAGACTCTTTGATTAGAGTGTGTCATGAGATAT 1865

seq\_name: gb\_sts:G18985

```

seq_documentation_block:
LOCUS      G18985          333 bp      DNA          STS          13-MAR-1996
DEFINITION Cow STS BMS2840.
ACCESSION  G18985
VERSION    G18985.1 GI:1223442
KEYWORDS   STS sequence; primer; sequence tagged site.
SOURCE     Bos taurus.
ORGANISM   Bos taurus

REFERENCE  1 (bases 1 to 333)
AUTHORS   Stone,R.T.
TITLE     Cow Markers
JOURNAL   Unpublished (1996)
COMMENT   Contact: Roger Stone
          U.S. Meat Animal Research Center
          U.S. Dept. of Agriculture - Agricultural Research Service
          P.O. Box 166, Clay Center, NE 68933
          Tel: (402) 762-4166
          Fax: (402) 762-4173

          Primer A: AGGACCCATAGGACGACAC
          Primer B: CCCTGGCAAGAGAAATTC
          STS size: 223.

FEATURES             Location/Qualifiers
     source           1..333
                     /organism="Bos taurus"
     STS              104..325
     primer_bind      104..123
     primer_bind      91 a 64 c 88 g 89 t 1 others
     BASE_COUNT       91 a 64 c 88 g 89 t 1 others
     ORIGIN

alignment_scores:
  Quality: 40.00      Length: 12
  Ratio: 3.333       Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 50.000

alignment_block:
US-08-653-294-17 x G18985/rev ..

Align seg 1/1 to reverse of: G18985 from: 1 to: 333

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||:
273 TATAGACTTTTGTAGTCAGAGAGCTCTCTTGAGATTC 238

seq_name: gb_ba2:APU96137

seq_documentation_block:
LOCUS      APU96137       7948 bp      DNA          BCT          22-DEC-1998
DEFINITION Anabaena PCC7120 apcABC gene cluster, phycobilisome core-membrane
          linker protein (apcE), allophycocyanin alpha subunit (apcA),
          allophycocyanin beta subunit (apcB) and phycobilisome core linker
          protein Lc7.8 (apcC) genes, complete cds.
ACCESSION  U96137
VERSION    U96137.1 GI:2072510
KEYWORDS
SOURCE     Anabaena PCC7120.
ORGANISM   Anabaena PCC7120.
REFERENCE  1 (bases 5592 to 7391)
AUTHORS   Cai,Y.A., Schwarts,S.H. and Glazer,A.N.
TITLE     Transposon insertion in genes coding for the biosynthesis of
          structural components of the Anabaena sp. phycobilisome
          Photosyn. Res. (1997) In press
REFERENCE  2 (bases 1 to 1560)
AUTHORS   Cai,Y.A., Lantoine,F.L. and Glazer,A.N.
TITLE     Characterization of the Anabaena transposable element ISan1 that
          belongs to the phylogenetically wide-spread Tc1/mariner superfamily
          Unpublished
JOURNAL   Unpublished
REFERENCE  3 (bases 1500 to 7948)

```

---

```

AUTHORS      Cai,Y.A., Lantoine,F.L. and Glazer,A.N.
TITLE        Characterization of Anabaena sp. PCC 7120 mutants defective in the
          large core-membrane linker protein of the light-harvesting
          phycobilisomes
JOURNAL      Unpublished
REFERENCE    4 (bases 1 to 7948)
AUTHORS      Cai,Y.A.
TITLE        Direct Submission
JOURNAL      Submitted (02-APR-1997) Department of Molecular and Cell Biology,
          University of California, Berkeley, CA 94720-3206, USA
FEATURES     Location/Qualifiers
             1..7948
             /organism="Anabaena PCC7120"
             /strain="PCC7120"
             /isolate="mutant SB12"
             /db_xref="taxon:1174"
             complement(1..268)
             /note="hypothetical"
             /codon_start=1
             /transl_table=11
             /protein_id="AAC97585.1"
             /db_xref="GI:2072511"
             /translation="MKHQLFCNIFEPYQWRLOEAOKEVESGFERLSQMSYSVRFA
             LAATILSQEDSQSILPILVKMSEIQLFEMDEKMTTEEKAILQR"
             /note="direct repeat A, copy 1"
             /rpt_type=direct
             271..283
             /note="inverted repeat A, copy 1"
             /rpt_type=Inverted
             287..297
             /note="inverted repeat A, copy 2"
             /rpt_type=Inverted
             294..314
             /note="direct repeat A, copy 2"
             /rpt_type=direct
             369..379
             /note="inverted repeat B, copy 1"
             /rpt_type=Inverted
             383..393
             /note="inverted repeat B, copy 2; transposable element
             ISan1"
             /rpt_type=Inverted
             398..1432
             /organism="Anabaena PCC7120"
             /db_xref="taxon:1174"
             /transposon="ISan1"
             398..413
             /note="ISan1 terminal inverted repeat, copy 1 (IRL)"
             /rpt_type=Inverted
             423..434
             /note="inverted repeat C, copy 1"
             /rpt_type=Inverted
             437..448
             /note="inverted repeat C, copy 2"
             /rpt_type=Inverted
             456..459
             /note="for orfA and orfAB"
             join(466..816,816..1421)
             /note="orfAB; -1 frame shift in the frame-shift window
             during translation produces the full-length fusion protein
             orfAB, presumably the active protein"
             /codon_start=1
             /transl_table=11
             /protein_id="AAC97588.1"
             /db_xref="GI:2072514"
             /translation="MKAYSIDFRQKIDFTYKGTGIGISORQLANKFCVSLGFIEKLLKQY
             RETASIAPIKVTQTPPKLNEQMKILEEIVKAKNDLTSEIRFLKERTGITIGIST
             VDRMLQRIEISLAKKTHLASEKTERVQLLRVQFVLQGLAENLVFLDEAGANLSL
             LRVSARKKGRAYGCRPHQHVSIIGAIKLVISQYSILGTSGLTFEAYISOK
             LVPKLWEACVIMDCNSIHKGGIEKIESAGAKLLVLPYSPDFSPICNWSKIKNF
             LRSTARSYPDLAKAIESAFNQVSLNDIYNWFTHSCYCHSPD"
             466..861

```



```

/note="orfA of ISan1"
/codon_start=1
/transl_table=11
/protein_id="AAC97586.1"
/db_xref="GI:2072512"
/translation="MKAYSLDFPROKIDFTYNTGTGTSORLANKFCVSLGFTFKLLKQY
RETASIPKVTIKOTPKLNEEQMKILEEIVKAKNDLTLSEIRFILKEKGTITIGIST
VDRMLORIEISLKKKHCTPOKRLKEFNY"
<810..1421
/note="orfB of ISan1; follows translation frame-shift
window"
/codon_start=1
/transl_table=11
/protein_id="AAC97587.1"
/db_xref="GI:2072513"
/translation="KKTILHASEKTERVOLLRVQFVLOLOGILAENLVFLDEAGANLS
LIRASAKKKRAYGRPQRKHQNVSIIGALIKGVISQISLGTSDGLTFEAYISQ
KLVPKLEAAACVIMDNCISIHKGDIKELIESAGAKLIYLPYPSPDFSPICNWSKIRN
FLRSIARSYPDLAKAIESAFNOVSLNDIYNWFTHSCYCTSPD"
818..831
/note="inverted repeat D, copy 1"
/rpt_type=direct
/rpt_type=inverted
841..855
/note="inverted repeat D, copy 2; stem-loop formation from
inverted repeat D facilitates the -1 frame shift during
translation"
/rpt_type=direct
/rpt_type=inverted
1417..1432
/note="ISan1 terminal inverted repeat, copy 2 (IRR)"
/rpt_type=direct
/rpt_type=inverted
1591..1602
/note="inverted repeat E, copy 1"
/rpt_type=direct
/rpt_type=inverted
1606..1619
/note="direct repeat B, copy 1"
/rpt_type=direct
1609..1620
/note="inverted repeat E, copy 2"
/rpt_type=direct
/rpt_type=inverted
1624..1637
/note="direct repeat B, copy 2"
/rpt_type=direct
1908..1911
/gene="apcB"
1908..5323
/gene="apcE"
1925..5323
/gene="apcB"
/codon_start=1
/transl_table=11
/product="phycobilisome core-membrane linker protein"
/protein_id="AAC97589.1"
/db_xref="GI:2072515"
/translation="MSVAKGSSVARPQLYQTLAVATITQAEQQDFLGRGELDELA
SYFASGAKLEIAQLLTENSEIIVSRAANRIFVGGSPMAFLERPEPELAAVGGGG
DVRESMLGTIVYVETRGLENLRSINTSPGTPPGFRPINIARYGSNNAKLSR
DLSEFLAYAVIAGDPNIIVNTRGLREIENACSGEATIVALQIKKASLSYFRK
DPDAEISYQMDVLITFEKAPTSNKLORPSGDOGLQLPOLYFSAERERKVMK
TGLSATSEKNEVKAAYRIQERDITRAYSISLESKVRNGDSIMKREFRVLAKSPL
YQKQYOFFINSRVIELAFRHILGSPSREEVQKYSIISNGGLPALVDALVDSAEY
SDYFEGTIVYRLGLGQAEQCRNWPQDLNFYSAPFRKVPQITTFAYADRLPDQ
HPYGSNDPIEIOGALFPKTRNPSPAPFGKDRILIHQPGINNGVSNPSARG
LAPGLGPKFKLDQPLTGKKAAGSAYKFSSTQAVIKATYLOVFGDVTVEGQR
LKQVEIKLENGESVRDVRALAKSLDFKLYTVPYVCKAIEVIERLLGRPTYGRQ
ENKFIADSKGLIYAVDALLDSETEFEGDVTYERILPAGVALQRLRVGTR
EDVANPEKQTPREVLGTAKENTQPDIFRINQGVTKQRETKVKFRVAGKDKAA
IKTILSAAYQETREITAPYATQNEFSWESKLGNEITVKEFEGLYSNLYLKEFY
TPYNTVRIELGTIKGRAPIDQAEIRKYNQILATQIGIRAFINALVNSQYENAFGE
DTPVRRFPTLPAAFNFTOKLYNOLTKQNDVVIPEFKPVOARIQSDKTPILAKAFA
DLAAQAKMDKSKPLFLTELGRSYNDGRGSVEGVGTRTRKPARIVELTNGIGQAEKQ
LVINATYQVLDVFGGVPDYRYRTELDSKLNGEISVREFREIASSEIYRFRFTYF
IPNTKITEFLRHLGRAPATQGEIRQINKLLADNGLRUAAVEAIVDSPEYSRYFGEV
VPYPRFPSPAGNYLGSVQAADLVKSSSLSPSTLTTPGRDGR"

```

```

repeat_region 5330..5337
/note="direct repeat C, copy 1"
/rpt_type=direct
5338..5344
/note="direct repeat C, copy 2"
/rpt_type=direct
5345..5351
/note="direct repeat C, copy 3"
/rpt_type=direct
5352..5358
/note="direct repeat C, copy 4"
/rpt_type=direct
5359..5382
/note="direct repeat D, copy 1"
/rpt_type=direct
5383..5401
/note="direct repeat E, copy 1"
/rpt_type=direct
5402..5415
/note="direct repeat E, copy 2"
/rpt_type=direct
5416..5429
/note="direct repeat E, copy 3"
/rpt_type=direct
5430..5450
/note="direct repeat D, copy 2"
/rpt_type=direct
5451..5462
/note="inverted repeat F, copy 1"
/rpt_type=inverted
5463..5477
/note="inverted repeat F, copy 2"
/rpt_type=inverted

alignment_scores:
  Quality: 40.00      Length: 12
  Ratio: 3.636       Gaps: 0
  Percent Similarity: 91.667   Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x APU96137/rev ..

Align seg 1/1 to reverse of: APU96137 from: 1 to: 7948

1 TYRArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||:|||||
126 TACAACACTCTTCTCAAGAGGATCAATCTCAGATAC 91

seq_name: gb_inl:CEF14B6

seq_documentation_block:
LOCUS CEF14B6 29283 bp DNA INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid F14B6, complete sequence.
ACCESSION Z81502
VERSION Z81502.1 GI:2814053
KEYWORDS HTG.
SOURCE
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE
AUTHORS
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans

```

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Nature 368 (6466), 32-38 (1994)  
94150718  
2 (bases 1 to 29283)  
White, S.  
Direct Submission  
Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rw@ematode.wustl.edu  
On Jan 28, 1998 this sequence version replaced gi:1665918.  
Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.  
For a graphical representation of this sequence and its analysis  
see:-  
http://webcace.sanger.ac.uk/cgi-  
bin/display?db=wormace&class=Sequence &object=F14B6

Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone F14B6.

It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.

The true left end of clone F41D3 is at 29177 in this sequence. The  
true right end of clone C15C6 is at 19321 in this sequence. The  
start of this sequence (1..109) overlaps with the end of sequence  
281034.

The end of this sequence (29177..29283) overlaps with the start of  
sequence 281537.

FEATURES  
source

Location/Qualifiers

1..29283

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="I"

/clone="F14B6"

/complement(947..2169)

/gene="F14B6.1"

/complement(join(947..1390,1835..2063,2114..2169))

/genes="F14B6.1"

/notes="cDNA EST yk242e8.5 comes from this gene; cDNA EST  
yk242e8.3 comes from this gene; cDNA EST yk414b10.5 comes  
from this gene; cDNA EST yk477b2.3 comes from this gene;  
cDNA EST yk477b2.5 comes from this gene; cDNA EST  
yk608e9.3 comes from this gene"

/codon\_start=1

/protein\_id="CAB04105.1"

/db\_xref="GI:3875879"

/db\_xref="SPTREMBL:O62161"

/translation="MPSVFELIIFGILNWKFTQPTPEPASRFFTRVCOQLRETV  
LIEDSGRVQFVSVGSGKPYSPHYGCAEQMWDVYGEQPHREEDQDQDTE  
GYEESEIEDSDQDITFTVKTPTAPLSPETENFEDKMKHFOATDILNLYLCKGT  
DNGIVNLRIQSYGVOLKOVSSGSKSKEIEGILKSLQKFEQIEGLNGEISEAE  
ENNESDISEFVENLEDSDFE"

/complement(2938..3345)

/genes="F14B6.7"

/complement(join(2938..3065,3115..3212,3260..3345))

/genes="F14B6.7"

/codon\_start=1

/protein\_id="CAB54218.1"

/db\_xref="GI:582443"

/db\_xref="SPTREMBL:O62165"

/translation="MKARKILITVAIALLSLAYRYYSORTMEFLIEEDQMLQREWEY  
SEFQCHSEILNVNFKFSILKFFKSNFNNIYKFTISSQNSTADYNSQVQISF  
SDVQTFWIIHPKISGPEILLVSRPDRSRNVIRHTWMSQENIKYLFILGLG  
ANMCKIKEVVMAKLFQDVTWVMEDRYKLSFKTLTLLFGVSKVPSAQILKID  
GDVLFEPFLSTIKNENMINVTSAVYKIAEAGVPIVSNCKRFLKYFNFRNSLI  
SFGCTRYAFYLAGPFLYTRSAALRLLEASKHRNFHKIEDTLITGLVADDDVSRVQL  
HRINLQKQKGDLLVFAWSPNDPEYKDLIYKTMSSQQQFKEKRLQLEINV"

/complement(4016..6657)

/genes="F14B6.2"

/complement(join(4016..4423,4517..5309,5485..5820,  
6093..6275,6563..6657))

/genes="F14B6.2"

/codon\_start=1  
/protein\_id="CAB04106.1"  
/db\_xref="GI:3875880"  
/db\_xref="SPTREMBL:O62162"  
/translation="MSTSSSVLPTPTPIGKQVIDLKAGAILRENARKSVLFFGGPLSV  
VLVLILKICISFIWIREHETDFRCAWNGGQPTDYMIRAKRSIAFERSDPRKFS  
KIPDDDDNSKCTIVINKDKKKKPLGRSPANDIISDAPKIDPRDFGLPPLNFQ  
RKDSNHIPAVPVPDDDLPMVTTLTQIIVQLSGEETGVVKMTIRNVRKVGK  
DEQFIEDIANAPRIYVSKIKSRGGEFEHFKLKKDEKEKKEKKKGGK  
LKAKKLEAEKLLNNSTSTSVIPDVTAVSRPKGKGLKSKKLNQPKSDEVTKQPR  
HLVGRSSLSPTPRPKSSWNGVLDPODPNFSNOQGIIRGYPMOKASKRSPATTASQ  
KDEDEEKLKLESTRMMDDELKTYTFLAPKTQETQKVEEGVYQIAEHASRQK  
PMLKTEFEATSTTEKPLPGMSKQDKWEKKEAFEPVPPISLDLQSPSHSGSPAQ  
APOHIFPESASTSTPAFHFIPPSSEAPYVVEINDADTETVYVSSVTPMYRPFEN  
MGHLESTKSYDPLHLEVSFESEFCVCCVQLLCHVFDP"  
complement(9869..9940)  
/genes="F14B6.t2"  
/note="CAT Met M-trna; predicted using tRNAscan-SE-1.1.1;  
preliminary prediction; similar to tRNA-Met"  
complement(9869..9940)  
/genes="F14B6.t2"  
13752..13823  
/genes="F14B6.t1"  
/note="CAT Met M-trna; predicted using tRNAscan-SE-1.1.1;  
preliminary prediction; similar to tRNA-Met"  
13752..13823  
/genes="F14B6.t1"  
15024..16895  
/genes="F14B6.3"  
/genes="F14B6.3"  
/note="predicted using Genefinder; cDNA EST yk329e12.5  
comes from this gene; cDNA EST EMEL:CO8202 comes from this  
gene; cDNA EST EMEL:CO9873 comes from this gene; cDNA EST  
yk198f11.3 comes from this gene; cDNA EST yk198f11.5 comes  
from this gene; cDNA EST yk329e12.3 comes from this gene;  
cDNA EST yk338a7.3 comes from this gene; cDNA EST  
yk338a7.5 comes from this gene; cDNA EST yk404g5.3 comes  
from this gene; cDNA EST yk404g5.5 comes from this gene;  
cDNA EST yk415a6.3 comes from this gene; cDNA EST  
yk415a6.5 comes from this gene; cDNA EST yk498b10.3 comes  
from this gene; cDNA EST yk509c5.3 comes from this gene;  
cDNA EST yk573a12.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB04104.1"  
/db\_xref="GI:3875878"  
/db\_xref="SPTREMBL:O62160"  
/translation="MLSELOFSSKSLGYEDYFKPIAPSSPRFFTKYCNQKKTIV  
DSNNIIRPFVDESNEREPLEVAHLGSECHMRNRMVGENIAENNETDEDETDYE  
SETSDSELGQSPNPNLSEALNCTIENFDQVDEVRHKDFHRSQRLVLEICLGNM  
NSFLQENLTVDDISTGACALLNETAEKLFKMFSSSGISGRKESGNTKLIKFGEL  
EIRGDEEDDDDESVEGEEDSDDEDDQEGESSIEPRLSETFKKSARESLSS  
EAPVKDVIYIKQFADREHLESENEPRMSDEAKLVEDLIEESDESDFIIGRDE  
ELDDNSGKEEELLEDVTDQEDSDEDDQFFSL"  
17155..19537  
/genes="F14B6.6"  
/join(17155..17218,17267..17455,17898..18232,18389..18591,  
19031..19178,19319..19537)  
/genes="F14B6.6"  
/note="cDNA EST yk549c2.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB04109.2"  
/db\_xref="GI:582444"  
/db\_xref="SPTREMBL:O62165"  
/translation="MKARKILITVAIALLSLAYRYYSORTMEFLIEEDQMLQREWEY  
SEFQCHSEILNVNFKFSILKFFKSNFNNIYKFTISSQNSTADYNSQVQISF  
SDVQTFWIIHPKISGPEILLVSRPDRSRNVIRHTWMSQENIKYLFILGLG  
ANMCKIKEVVMAKLFQDVTWVMEDRYKLSFKTLTLLFGVSKVPSAQILKID  
GDVLFEPFLSTIKNENMINVTSAVYKIAEAGVPIVSNCKRFLKYFNFRNSLI  
SFGCTRYAFYLAGPFLYTRSAALRLLEASKHRNFHKIEDTLITGLVADDDVSRVQL  
HRINLQKQKGDLLVFAWSPNDPEYKDLIYKTMSSQQQFKEKRLQLEINV"

gene

```
CDS
Join(20434..20524,21137..21250,21298..21659,21969..22168,
22218..22344,22393..22482,22538..22684)
/gene="F14B6.4"
/codon_start=1
/protein_id="CAB04107.1"
/db_xref="GI:3875881"
/db_xref="SPTREMBL:O62163"
/translation="MEISLQKLFLEFLFANASSVSGGSGFYLLKLSIGLLNYQID
KNLOKEHEHKLALNCKMSGRIFNSVPETKFGSSFTISFADIKHTSWLYLPKFO
NMSSEIIMIVASRTDSPARNVLRKTWNPENSEIKDGRMKALFLVGMDDGDSRM
RKVNEEARIVDMVVDLKDYYELPEKSLITLLYGTSKASEFKLIGKIDEDINFFP
DKILPELLONLDPSSSIYGMFAEGGVYVDKEHRWFVPDSTTGCDFMFPYTGGLF
YLVTOAARKILNATKRIPIEDALINGILANDCKPRIHLPEIYDGHSLTDEDT
RGVLAWHTGKNDQOFLDYFYRKLSERLKKILDEKNKHFY"
23603..26251
/gene="F14B6.5"
Join(23603..23727,23784..24048,24104..24209,24287..24373,
24425..24635,24799..24965,25030..25343,25388..25514,
25566..25680,25729..25933,25988..26251)
/gene="F14B6.5"
/codon_start=1
/protein_id="CAB04108.1"
/db_xref="GI:3875882"
/db_xref="SPTREMBL:O62164"
/translation="MSVAAKPKRLDQIRGLAILSVLGFHYFPALFNGYLVGDQFF
VLSGLMCKLRAENETCTILFYSKRFKRIPLYLIIILSMISLYITFPDTSI
ETNOKSAMRALFVSNAPKSVQDDYFTMLARAVDIFTWLSLSEVQYFVLPVFLF
AKILSGLLSAFFRTPSTNSFVSFARWQFLGIMVIFLLSSHKSPKAEKNIIEEC
NKLIDEENNTQKAKPSILHLISYFPLISILICILYQNTDFVRLVYLAGALM
IVFENLILSKNLIYIGDTSYLIHWPYIYKWLAFDGVNLLILSLSSILAI
IVFEFEKWKYKLSISGLTIVLFLFLLVILINKDELIPDRENNHNSMNTDDAI

alignment_scores:
  Quality: 40.00      Length: 12
  Ratio: 83.333      Gaps: 0
Percent Similarity: 83.333      Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x CEF14B6
..
Align seg 1/1 to: CEF14B6 from: 1 to: 29283

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
|||||
3030 TACCGACTACTATCGCTAGAGTATTATCAATAT 3065

seq_name: gb_in2:AC005834

seq_documentation_block:
LOCUS AC005834 207672 bp DNA INV 21-OCT-1998
DEFINITION Drosophila melanogaster DNA sequence (P1s DS01589 (D279), DS01529 (D280), and DS02501 (D281)), complete sequence.
ACCESSION AC005834 AC004577 AC005733 AC004717 AC004718
VERSION AC005834.1 GI:3776556
KEYWORDS HTG.
SOURCE
Drosophila melanogaster (Subclones in sac from P1 clones DS01589 (D279), DS01529 (D280), and DS02501 (D281)) DNA.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 207672)
AUTHORS
Celniker,S.E., Harris,N.L., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Blazej,R.G., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 2L, region 28B1-28B4
Unpublished (1997)
TITLE
Sequencing of Drosophila chromosome 2L, region 28B1-28B4
JOURNAL
REFERENCE
2 (bases 1 to 207672)
AUTHORS
Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,

Svirskas,R.R., Harris,N.L., Aqbayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
Submitted (20-OCT-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to bdg@fruitfly.berkeley.edu.
Library locations: 53-17, 90-16, 5-27.
FEATURES
  Location/Qualifiers
    1..207672
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /chromosome="2L"
      /map="28B1-28B4"
      /clone="P1s DS01589 (D279), DS01529 (D280), and DS02501 (D281)"
      /note="These three p1s were completed as a project. The
coordinates are p1 end to p1 end. DS02501 (D281) extends
from bp 1 to bp 85,627. DS01529 (D280) extends from bp
74,691 to bp 155,821. DS01589 (D279) extends from bp
126,132 to bp 207,672."
BASE COUNT 61698 a 42102 c 41711 g 62161 t
ORIGIN

alignment_scores:
  Quality: 40.00      Length: 12
  Ratio: 3.333      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 41.667

alignment_block:
US-08-653-294-17 x AC005834
..
Align seg 1/1 to: AC005834 from: 1 to: 207672

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
|||||
39804 TTTCGAATCATATATAAAGGGTTGCAATTCATAT 39839

seq_name: gb_htg1:HSB348B13

seq_documentation_block:
LOCUS HSB348B13 229644 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens chromosome X clone RP13-348B13, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL121872
VERSION AL121872.2 GI:6249374
KEYWORDS HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 229644)
AUTHORS
Pavitt,R.
TITLE
Direct Submission
JOURNAL
Submitted (03-NOV-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquy@esanger.ac.uk Clone requests: clonerequest@esanger.ac.uk.
On Nov 4, 1999 this sequence version replaced gi:6065938.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
```

sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: b3348B13 Contig\_ID: 00058 acc-AL121872 Length: 1125 bp Unfinished: b3348B13 Contig\_ID: 00082 acc-AL121872 Length: 2320 bp Unfinished: b3348B13 Contig\_ID: 00093 acc-AL121872 Length: 2597 bp Unfinished: b3348B13 Contig\_ID: 00119 acc-AL121872 Length: 1185 bp Unfinished: b3348B13 Contig\_ID: 00121 acc-AL121872 Length: 3814 bp Unfinished: b3348B13 Contig\_ID: 00131 acc-AL121872 Length: 2607 bp Unfinished: b3348B13 Contig\_ID: 00137 acc-AL121872 Length: 3158 bp Unfinished: b3348B13 Contig\_ID: 00180 acc-AL121872 Length: 7824 bp Unfinished: b3348B13 Contig\_ID: 00198 acc-AL121872 Length: 1489 bp Unfinished: b3348B13 Contig\_ID: 00304 acc-AL121872 Length: 2018 bp Unfinished: b3348B13 Contig\_ID: 00306 acc-AL121872 Length: 2278 bp Unfinished: b3348B13 Contig\_ID: 00319 acc-AL121872 Length: 1043 bp Unfinished: b3348B13 Contig\_ID: 00340 acc-AL121872 Length: 1556 bp Unfinished: b3348B13 Contig\_ID: 00346 acc-AL121872 Length: 1602 bp Unfinished: b3348B13 Contig\_ID: 00349 acc-AL121872 Length: 1077 bp Unfinished: b3348B13 Contig\_ID: 00386 acc-AL121872 Length: 2564 bp Unfinished: b3348B13 Contig\_ID: 00443 acc-AL121872 Length: 1526 bp Unfinished: b3348B13 Contig\_ID: 00447 acc-AL121872 Length: 2762 bp Unfinished: b3348B13 Contig\_ID: 00458 acc-AL121872 Length: 1355 bp Unfinished: b3348B13 Contig\_ID: 00471 acc-AL121872 Length: 1151 bp Unfinished: b3348B13 Contig\_ID: 00480 acc-AL121872 Length: 1365 bp Unfinished: b3348B13 Contig\_ID: 00495 acc-AL121872 Length: 4089 bp Unfinished: b3348B13 Contig\_ID: 00546 acc-AL121872 Length: 7184 bp Unfinished: b3348B13 Contig\_ID: 00565 acc-AL121872 Length: 8596 bp Unfinished: b3348B13 Contig\_ID: 00590 acc-AL121872 Length: 1569 bp Unfinished: b3348B13 Contig\_ID: 00595 acc-AL121872 Length: 4358 bp Unfinished: b3348B13 Contig\_ID: 00631 acc-AL121872 Length: 3299 bp Unfinished: b3348B13 Contig\_ID: 00652 acc-AL121872 Length: 6958 bp Unfinished: b3348B13 Contig\_ID: 00658 acc-AL121872 Length: 1568 bp Unfinished: b3348B13 Contig\_ID: 00677 acc-AL121872 Length: 1105 bp Unfinished: b3348B13 Contig\_ID: 00685 acc-AL121872 Length: 2677 bp Unfinished: b3348B13 Contig\_ID: 00723 acc-AL121872 Length: 2023 bp Unfinished: b3348B13 Contig\_ID: 00758 acc-AL121872 Length: 7514 bp Unfinished: b3348B13 Contig\_ID: 00776 acc-AL121872 Length: 4420 bp Unfinished: b3348B13 Contig\_ID: 00799 acc-AL121872 Length: 3158 bp Unfinished: b3348B13 Contig\_ID: 00812 acc-AL121872 Length: 1014 bp Unfinished: b3348B13 Contig\_ID: 00845 acc-AL121872 Length: 1660 bp Unfinished: b3348B13 Contig\_ID: 00882 acc-AL121872 Length: 6089 bp Unfinished: b3348B13 Contig\_ID: 00901 acc-AL121872 Length: 2425 bp Unfinished: b3348B13 Contig\_ID: 00921 acc-AL121872 Length: 1300 bp Unfinished: b3348B13 Contig\_ID: 00943 acc-AL121872 Length: 1499 bp Unfinished: b3348B13 Contig\_ID: 00976 acc-AL121872 Length: 2534 bp Unfinished: b3348B13 Contig\_ID: 00988 acc-AL121872 Length: 1480 bp Unfinished: b3348B13 Contig\_ID: 00998 acc-AL121872 Length: 1047 bp Unfinished: b3348B13 Contig\_ID: 01049 acc-AL121872 Length: 7401 bp Unfinished: b3348B13 Contig\_ID: 01055 acc-AL121872 Length: 1280 bp Unfinished: b3348B13 Contig\_ID: 01062 acc-AL121872 Length: 1437 bp Unfinished: b3348B13 Contig\_ID: 01063 acc-AL121872 Length: 1961 bp Unfinished: b3348B13 Contig\_ID: 01088 acc-AL121872 Length: 3475 bp Unfinished: b3348B13 Contig\_ID: 01090 acc-AL121872 Length: 2284 bp Unfinished: b3348B13 Contig\_ID: 01147 acc-AL121872 Length: 2193 bp Unfinished: b3348B13 Contig\_ID: 01171 acc-AL121872 Length: 1655 bp Unfinished: b3348B13 Contig\_ID: 01204 acc-AL121872 Length: 4332 bp Unfinished: b3348B13 Contig\_ID: 01240 acc-AL121872 Length: 5408 bp Unfinished: b3348B13 Contig\_ID: 01265 acc-AL121872 Length: 1757 bp Unfinished: b3348B13 Contig\_ID: 01283 acc-AL121872 Length: 5191 bp Unfinished: b3348B13 Contig\_ID: 01296 acc-AL121872 Length: 3846 bp Unfinished: b3348B13 Contig\_ID: 01341 acc-AL121872 Length: 1067 bp Unfinished: b3348B13 Contig\_ID: 01347 acc-AL121872 Length: 11938 bp Unfinished: b3348B13 Contig\_ID: 01351 acc-AL121872 Length: 3337 bp.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

#### FEATURES

source Location/Qualifiers  
1. .229644  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP13-348B13"  
/clone\_lib="RPC1-13.2"

BASE COUNT 60513 a 31906 c 32427 g 57543 t 47255 others  
ORIGIN

#### alignment\_scores:

Quality: 40.00 Length: 10  
Ratio: 4.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 70.000

#### alignment\_block:

US-08-653-294-17 x HSB348B13 ..

Align seg 1/1 to: HSB348B13 from: 1 to: 229644

3 LeuLeuIleArgArgIleAlaLeuArgTyr 12

126319 ATCTTAGTTCGAGAAATGCTGGAATAT 126348

seq\_name: gb\_htgl:CEY47H10

#### seq\_documentation\_block:

LOCUS CEY47H10 296589 bp DNA HTG 07-SEP-1999  
DEFINITION Caenorhabditis elegans chromosome I clone Y47H10, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in unordered pieces.  
ACCESSION 295311  
VERSION 295311.10 GI:5725202  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 296589)  
AUTHORS McWay, K.  
TITLE Direct Submission  
JOURNAL Submitted (07-SEP-1999) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwnematode.wustl.edu  
On Aug 10, 1999 this sequence version replaced gi:4938508.  
Order of segments is not known; 800 n's separate segments.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

#### FEATURES

source Location/Qualifiers  
1. .296589  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="I"  
/clone="Y47H10"

BASE COUNT 92991 a 55009 c 53606 g 92582 t 2401 others  
ORIGIN

#### alignment\_scores:

Quality: 40.00 Length: 12  
Ratio: 4.000 Gaps: 0

alignment_scores:	
Quality:	39.00
Ratio:	3.545
Percent Similarity:	91.667
Percent Identity:	58.333
Gaps:	0
Length:	12

**THIS PAGE BLANK (USPTO)**

OM of: US-08-653-294-17 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Feb 8, 2000 1:28 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cnll1/USPIO.spool/US08653294/runat\_04022000\_160701\_15807/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -QMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -SPART=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

#### Search information block:

Query: US-08-653-294-17  
Query length: 12  
Database: N\_Geneseq\_36:\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 590.520000

#### score\_list:

Sequence	Strtd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X13540	+	38.00	110.34	151.66	2469	Enterococcus faecalis genome c
N_Geneseq_36:X88170	+	36.00	120.91	391.11	321	EST clone DT478. New polynucleo
N_Geneseq_36:X780665	+	36.00	102.39	420.45	2715	Human DNAX toll-like receptor
N_Geneseq_36:X79140	+	36.00	91.54	1.7e+03	9487	Hepatitis C virus gene HC-G9 c
N_Geneseq_36:X781920	+	35.00	127.63	16.50	98	Interferon-gamma binding oligon
N_Geneseq_36:X78465	+	35.00	124.42	24.93	142	Lactococcus lactis pl promoter
N_Geneseq_36:X708093	+	35.00	102.25	428.18	1830	Chloroperoxidase gene. Enzymat
N_Geneseq_36:X730875	+	35.00	101.10	495.84	2088	Chloroperoxidase cDNA. Use of
N_Geneseq_36:X13205	+	35.00	100.22	555.09	2311	Enterococcus faecalis genome c
N_Geneseq_36:X75954	+	35.00	100.11	562.84	2340	S. cerevisiae scaur1R gene. Ne
N_Geneseq_36:X75955	+	35.00	100.11	562.84	2340	S. cerevisiae scaur1S gene. Ne
N_Geneseq_36:X709350	+	35.00	100.11	562.84	2340	Aureobasidin sensitivity gene
N_Geneseq_36:X709355	+	35.00	100.11	562.84	2340	Aureobasidin resistance gene s
N_Geneseq_36:X20864	+	35.00	99.44	613.63	2529	Polynucleotide sequence from t
N_Geneseq_36:X751612	+	35.00	98.49	693.21	2822	Cervularia verruculosa haloper
N_Geneseq_36:X07548	+	35.00	93.16	1.4e+03	5216	PKS90 TIR coupling vector. New
N_Geneseq_36:X07549	+	35.00	93.14	1.4e+03	5230	PTREX1 mutant beta-toxin expre
N_Geneseq_36:X07550	+	35.00	93.14	1.4e+03	5231	PTREX1A mutant beta-toxin expre
N_Geneseq_36:X74441	+	35.00	92.06	1.6e+03	5924	Staphylococcus aureus contig s
N_Geneseq_36:X58840.5	+	35.00	69.47	2.8e+04	80073	Continuation (6 of 6) of T58
N_Geneseq_36:X20573	+	34.50	85.76	3.5e+03	9974	Polynucleotide sequence from t
N_Geneseq_36:X19288	+	34.00	118.99	50.00	176	Human gene signature HUMGS00305
N_Geneseq_36:X726331	+	34.00	112.27	118.38	382	Human gene signature HUMGS08571
N_Geneseq_36:X21079	+	34.00	111.18	136.09	433	Polynucleotide sequence from th
N_Geneseq_36:X784032	+	34.00	105.08	297.62	875	DNA encoding a Staphylococcus a
N_Geneseq_36:X53425	+	34.00	105.08	297.62	875	DNA encoding a Staphylococcus a
N_Geneseq_36:X13210	+	34.00	101.56	467.83	1314	Enterococcus faecalis genome c
N_Geneseq_36:X12200	+	34.00	99.04	646.31	1757	Human neuronal nicotinic acety
N_Geneseq_36:T41477	+	34.00	97.75	761.85	2037	Haemophilus adhesion protein H
N_Geneseq_36:T60336	+	34.00	96.89	850.97	2250	TIM 01 mutant coding sequence.
N_Geneseq_36:X60302	+	34.00	94.96	1.1e+03	2812	Lucilia cuprina ecdysteroid re
N_Geneseq_36:X60335	+	34.00	93.39	1.3e+03	3369	TIM coding sequence. Nuclear t
N_Geneseq_36:X27128	+	34.00	93.39	1.3e+03	3369	Drosophila timeless gene prote
N_Geneseq_36:X087444	+	34.00	92.91	1.4e+03	3560	Drosophila semaphorin I cDNA.
N_Geneseq_36:X40189	+	34.00	92.82	1.4e+03	3599	Lung cancer associated gene. N
N_Geneseq_36:X40192	+	34.00	92.49	1.5e+03	3736	Lung cancer associated gene. N
N_Geneseq_36:X40191	+	34.00	92.33	1.5e+03	3807	Lung cancer associated gene. N
N_Geneseq_36:X40190	+	34.00	91.77	1.6e+03	4060	Lung cancer associated gene. N
N_Geneseq_36:X13357	+	34.00	91.66	1.7e+03	4114	Enterococcus faecalis genome c
N_Geneseq_36:T60334	+	34.00	91.54	1.7e+03	4170	TIM splice variant coding sequ
N_Geneseq_36:X27129	+	34.00	91.54	1.7e+03	4170	Drosophila timeless gene prote

N\_Geneseq\_36:X20635 - 34.00 90.99 1.8e+03 4444 ! Polynucleotide sequence fro  
N\_Geneseq\_36:T60333 - 34.00 89.63 2.2e+03 5198 ! TIM gene. Nuclear transloca  
N\_Geneseq\_36:X27127 - 34.00 89.63 2.2e+03 5198 ! Timeless gene. New protein  
N\_Geneseq\_36:X13141 + 34.00 87.25 2.9e+03 6835 ! Enterococcus faecalis genom

seq\_name: N\_Geneseq\_36:X13540

seq\_documentation\_block:  
ID X13540 standard; DNA; 2469 BP.

AC X13540;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:603.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; U08985.

PR 14-NOV-1997; US-066009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

PS Claim 1; Page 1853-1855; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

SQ Sequence 2469 BP; 771 A; 471 C; 483 G; 738 T;

#### alignment\_scores:

Quality: 38.00 Length: 12  
Ratio: 3.455 Gaps: 0  
Percent Similarity: 91.667 Percent Identity: 50.000

#### alignment\_block:

US-08-653-294-17 x X13540/rev ..

Align seg 1/1 to reverse of: X13540 from: 1 to: 2469

1 TyrArgLeuLeuLeuArgArgIleAlaLeuArgTyr 12

|||||

1322 TTCAGCTTCTCTCAACCGGGTTTCGTCGGTAT 1287

seq\_name: N\_Geneseq\_36:X88170

#### seq\_documentation\_block:

ID V88170 standard; cDNA; 321 BP.

AC V88170;

DT 12-FEB-1999 (first entry)

DE EST clone DT478.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

OS Homo sapiens.

PN W09845437-A2.





US-08-653-294-17 x Q81920



1514 TTAGGATAATCTTCAGGAGATAGCCCTTTTAT 1549

seq\_name: N\_Geneseq\_36:Q75954

seq\_documentation\_block:

ID Q75954 standard; DNA; 2340 BP.

AC Q75954;

DT 04-AUG-1995 (first entry)

DE S. cerevisiae scaur1R gene.

KW Aureobasidin-resistance; aureobasidin-sensitivity; scaur1R gene;

KW mycosis; antimycotic; fungicide; diagnosis; therapy; ds.

OS Saccharomyces cerevisiae DKD5D.

FH Key Location/Qualifiers

FT misc\_difference 765..766

FT /\*tag= a

FT /note= "bases at positions 765-66 are unclear

FT in the specification"

FT misc\_difference 1544

FT /\*tag= b

FT /note= "base at position 1544 is unclear in the

FT specification"

PN AU9463129-A.

PD 01-DEC-1994.

PF 16-MAY-1994; 063129.

PR 24-MAY-1993; JP-142523.

PR 28-DEC-1993; JP-348893.

PA (TAKI ) TAKARA SHUZO CO LTD.

PI Kato I, Okada T, Takesako K;

DR WPI: 95-023286/04.

DR P-PSDB: R67689.

PT New genes regulating aureobasidin sensitivity - used to develop

PT prods. for the diagnosis and treatment of diseases such as

PT mycoses

PS Disclosure: Page 67-68: 110pp; English.

CC Saccharomyces cerevisiae DKD5D, which is sensitive to aureobasidin,

CC was mutagenized with EMS and genomic libraries of 2 resistant

CC strains were prepared. Resistant gene scaur1R (Q75954) was isolated

CC from 1 library, and resistant gene scaur2R (Q75956) from the other.

CC A DNA fragment of scaur1 was used as a probe to isolate a

CC sensitive gene, scaur1S (Q75955), from the sensitive strain. A

CC T852A mutation leading to a F158Y conversion gave rise to the

CC resistance of scaur1R. 584 A; 461 C; 429 G; 863 T;

SQ Sequence 2340 BP; 584 A; 461 C; 429 G; 863 T;

alignment\_scores:

Quality: 35.00 Length: 10

Ratio: 3.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 50.000

alignment\_block:

US-08-653-294-17 x Q75954

Align seg 1/1 to: Q75954 from: 1 to: 2340

3 LeuLeuIeaArgArgIleAlaLeuArgTyr 12

:::::::::::|||||

2291 ATCTAGTGAAGAAGGGTGGTTTGCCTTAT 2320

seq\_name: N\_Geneseq\_36:Q75955

seq\_documentation\_block:

ID Q75955 standard; DNA; 2340 BP.

AC Q75955;

DT 04-AUG-1995 (first entry)

DE S. cerevisiae scaur1S gene.

KW Aureobasidin-resistance; aureobasidin-sensitivity; scaur1S gene;

KW mycosis; antimycotic; fungicide; diagnosis; therapy; ds.

OS Saccharomyces cerevisiae DKD5D.

PN AU9463129-A.

PD 01-DEC-1994.

PF 16-MAY-1994; 063129.

PR 24-MAY-1993; JP-142523.

PR 28-DEC-1993; JP-348893.

PA (TAKI ) TAKARA SHUZO CO LTD.

PI Kato I, Okada T, Takesako K;

DR WPI: 95-023286/04.

DR P-PSDB: R67690.

PT New genes regulating aureobasidin sensitivity - used to develop

PT prods. for the diagnosis and treatment of diseases such as

PT mycoses

PS Disclosure: Page 70-72: 110pp; English.

CC Saccharomyces cerevisiae DKD5D, which is sensitive to aureobasidin,

CC was mutagenized with EMS and genomic libraries of 2 resistant

CC strains were prepared. Resistant gene scaur1R (Q75954) was isolated

CC from 1 library, and resistant gene scaur2R (Q75956) from the other.

CC A DNA fragment of scaur1 was used as a probe to isolate a

CC sensitive gene, scaur1S (Q75955), from the sensitive strain. A

CC T852A mutation leading to a F158Y conversion gave rise to the

CC resistance of scaur1R. 583 A; 462 C; 430 G; 865 T;

SQ Sequence 2340 BP; 583 A; 462 C; 430 G; 865 T;

alignment\_scores:

Quality: 35.00 Length: 10

Ratio: 3.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 50.000

alignment\_block:

US-08-653-294-17 x Q75955

Align seg 1/1 to: Q75955 from: 1 to: 2340

3 LeuLeuIeaArgArgIleAlaLeuArgTyr 12

:::::::::::|||||

2291 ATCTAGTGAAGAAGGGTGGTTTGCCTTAT 2320

seq\_name: N\_Geneseq\_36:Q75950

seq\_documentation\_block:

ID T09350 standard; DNA; 2340 BP.

AC T09350;

DT 29-MAR-1996 (first entry)

DE Aureobasidin sensitivity gene scaur1S.

KW Aureobasidin resistance; selectable marker; fungus transformation;

KW Chromosome integration vector; scaur1S; ds.

OS Saccharomyces cerevisiae.

FH Key Location/Qualifiers

FT cds

FT 380..1585

FT /\*tag= a

PN EP-692534-A2.

PD 17-JAN-1996.

PF 15-JUN-1995; 304165.

PR 29-JUN-1994; JP-168611.

PR 30-MAR-1995; JP-095955.

PR 30-MAR-1995; JP-095831.

PR 17-MAY-1995; JP-141393.

PA (TAKI ) TAKARA SHUZO CO LTD.

PI Kato I, Ogawa J, Okada T, Oshima A, Takesako K;

PI Yoshioka H;

DR WPI: 96-070176/08.

DR P-PSDB: R88127.

PT Fungal chromosome integration vector contg. aureobasidin resistance

PT gene - which acts as selection marker, useful in genetic

PT engineering of fungi

PS Disclosure: Page 26-28: 66pp; English.

CC The wild-type aureobasidin sensitivity gene scaur1SR (T09350) was

CC obtained from a genomic library of Saccharomyces cerevisiae using a

CC mutant aureobasidin resistance gene (T09355) as probe. The gene

CC encodes a protein (R88127) sensitive to aureobasidin.

CC Mutagenesis of the gene at codon 240 resulting in a substitution,

CC deletion or insertion at Ala-240 is used to impart aureobasidin

CC resistance, and the mutated DNA may be used as a selectable

CC marker useful in fungus transformation.

CC Sequence 2340 BP; 583 A; 481 C; 411 G; 865 T;

```

alignment_scores:
  AC      Quality: 35.00      Length: 10
  DE      Ratio: 3.500       Gaps: 0
  KW      Percent Similarity: 100.000      Percent Identity: 50.000

alignment_block:
  US-08-653-294-17 x T09350      ..
  Align seg 1/1 to: T09350      from: 1 to: 2340

      3 LeuLeuIleArgArgIleAlaLeuArgTyr 12
      ::::|||||:|||||:|||||:|||||:
  2291 ATCTAGTAAAGGCGTGGTTTGGCTTAT 2320

seq_name: N_Geneseq_36:T09355

seq_documentation_block:
  ID T09355 standard; DNA; 2340 BP.
  AC T09355;
  DT 29-MAR-1996 (first entry)
  DE Aureobasidin resistance gene scau1R.
  KW Aureobasidin resistance; selectable marker; fungus transformation;
  KW chromosome integration vector; Saccharomyces cerevisiae; scau1R;
  KW ds.
  OS Synthetic.
  FH Key
  FT Cds
  FT EP-692534-A2.
  PN 17-JAN-1996.
  PR 15-JUN-1995; 304165.
  PR 29-JUN-1994; JP-168611.
  PR 30-MAR-1995; JP-085955.
  PR 30-MAR-1995; JP-095831.
  PR 17-MAY-1995; JP-141393.
  PA (TAKI ) TAKARA SHUZO CO LTD.
  PI Kato I, Ogawa J, Okado T, Oshima A, Takesako K;
  PI Yoshioka H;
  DR WPI: 96-070176/08.
  DR P-PSDB; R88132.
  PT Fungal chromosome integration vector contg. aureobasidin resistance
  PT gene - which acts as selection marker, useful in genetic
  PT engineering of fungi
  PS Disclosure; Page 23-24; 66pp; English.
  CC The aureobasidin resistance gene scau1R (T09355) was obt'd. from
  CC a genomic library of Saccharomyces cerevisiae following EMS
  CC mutagenesis. Comparison with the wild-type aureobasidin
  CC sensitive gene (T09350) showed a T to A mutation at position 852,
  CC resulting in a Fl58Y mutation in the encoded protein (R88132).
  CC The gene may be used as a selectable marker component of chromosome
  CC integration vectors used for fungus transformation.
  SQ Sequence 2340 BP; 584 A; 505 C; 387 G; 864 T;

alignment_scores:
  AC      Quality: 35.00      Length: 10
  DE      Ratio: 3.500       Gaps: 0
  KW      Percent Similarity: 100.000      Percent Identity: 50.000

alignment_block:
  US-08-653-294-17 x T09355      ..
  Align seg 1/1 to: T09355      from: 1 to: 2340

      3 LeuLeuIleArgArgIleAlaLeuArgTyr 12
      ::::|||||:|||||:|||||:|||||:
  2291 ATCTAGTAAAGGCGTGGTTTGGCTTAT 2320

seq_name: N_Geneseq_36:X20664

seq_documentation_block:
  ID X20664 standard; DNA; 2529 BP.

```

```

AC X20664;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
  enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI: 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 825-826; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 2529 BP; 589 A; 670 C; 699 G; 567 T;

alignment_scores:
  AC      Quality: 35.00      Length: 11
  DE      Ratio: 4.375       Gaps: 0
  KW      Percent Similarity: 72.727      Percent Identity: 72.727

alignment_block:
  US-08-653-294-17 x X20664/rev      ..
  Align seg 1/1 to reverse of: X20664      from: 1 to: 2529

      1 TyrArgLeuLeuIleArgArgIleAlaLeuArg 11
      :|||||:|||||:|||||:|||||:
  2398 TACAGATTATACGCCGATCGATTGCGTTACGC 2366

seq_name: N_Geneseq_36:T51612

seq_documentation_block:
  ID T51612 standard; DNA; 2822 BP.
  AC T51612;
  DT 30-APR-1997 (first entry)
  DE Curvularia verruculosa haloperoxidase gene.
  KW Haloperoxidase; halide oxidation; halogenation; antimicrobial;
  KW disinfectant; ss.
  OS Curvularia verruculosa strain CBS 147.63.
  FH Key
  FT Cds
  FT 477..2279
  FT Location/Qualifiers
  FT Cds
  FT primer_bind
  FT complement (477..496)
  FT /tag= a
  FT /tag= b
  FT /note= "primer aHap1 binding site"
  FT primer_bind
  FT 2293..2311
  FT /tag= c
  FT /note= "primer aHap1A binding site"
  FT
  PN WO9704102-A1.
  PD 06-FEB-1997.
  PR 09-JUL-1996; U11458.
  PR 14-JUL-1995; US-001194.
  PR 21-FEB-1996; US-603534.
  PA (NOVO ) NOVO NORDISK BIOTECH INC.
  PA (NOVO ) NOVO NORDISK AS.
  PI Berka RW, Cherry J, Fuglsang C, Halkier T, Oxenboll KM;
  DR WPI: 97-132641/12.
  DR P-PSDB; W12042.
  PT Haloperoxidase from Curvularia verruculosa - useful for oxidn. of
  PT halide, halogenation and, in presence of hydrogen peroxide and
  PT thiocyanate, as antimicrobial
  PT Claim 28; Page 33-35; 58pp; English.

```



**THIS PAGE BLANK (USPTO)**

Sequence	Strd Orig	zScore	EScore	Len	Documentation
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-468-589-3 +		37.00	113.18	69.86	1756
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-700-636-3 +		37.00	113.18	69.86	1756
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-210-222-8 +		35.00	132.72	5.70	98
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-679-405-1 +		35.00	101.15	326.65	2822
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-842-799-1 +		35.00	101.15	326.65	2822
/cgnl_6/pdata/2/ina/PCtUS9 COMB. seq:PGT-US96-11458-1 +		35.00	101.15	326.65	2822
/cgnl_6/pdata/2/ina/PCtUS9 COMB. seq:PGT-US96-05320A-897 +		34.00	96.50	593.42	3333
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-619-198-2 +		34.00	95.71	657.02	3369
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-121-713D-59 +		34.00	95.19	702.16	3560
/cgnl_6/pdata/2/ina/3B COMB. seq:US-08-833-268-59 +		34.00	95.19	702.16	3560
/cgnl_6/pdata/2/ina/5D COMB. seq:US-09-060-692-59 +		34.00	95.19	702.16	3560
/cgnl_6/pdata/2/ina/PCtUS9 COMB. seq:PGT-US94-10351A-59 +		34.00	94.59	702.16	3560
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-619-198-4 +		34.00	93.70	849.55	4170
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-619-198-8 +		34.00	91.64	1.1e+03	5192
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-619-198-1 +		34.00	91.63	1.1e+03	5198
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-038-682-6 +		34.00	86.14	2.2e+03	9323
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-302-832-6 +		34.00	86.14	2.2e+03	9323
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-530-198-6 +		34.00	86.14	2.2e+03	9323
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-469-880-6 +		34.00	86.14	2.2e+03	9323
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-728-470-6 +		34.00	86.14	2.2e+03	9323
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-617-697-6 +		34.00	86.14	2.2e+03	9323
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-558-735-9 +		33.00	113.41	67.93	342
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-558-735-6 +		33.00	113.33	68.85	345
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-833-659A-11 +		33.00	112.77	73.60	366
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-558-735-7 +		33.00	111.04	91.89	440
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-558-735-7 +		33.00	109.44	112.89	522
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-186-833-1 +		33.00	101.99	293.31	1153
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-997-080-193 +		33.00	101.90	296.68	1164
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-997-362-193 +		33.00	101.90	296.68	1164
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-159-340A-1 +		33.00	92.04	1.1e+03	3328
/cgnl_6/pdata/2/ina/5B COMB. seq:US-08-368-803-6 +		33.00	92.01	1.1e+03	3336
/cgnl_6/pdata/2/ina/backfile.sesl seq:522324-11 +		33.00	91.85	1.1e+03	3395
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-742-026-1 +		33.00	90.99	1.2e+03	3720
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-742-026-22 +		33.00	90.99	1.2e+03	3720
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-663-566A-3 -		33.00	87.44	1.9e+03	5426
/cgnl_6/pdata/2/ina/5C COMB. seq:US-08-023-610-3 -		33.00	87.44	1.9e+03	5426
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-288-065A-3 -		33.00	87.44	1.9e+03	5426
/cgnl_6/pdata/2/ina/5D COMB. seq:US-08-362-240A-3 -		33.00	87.44	1.9e+03	5426
/cgnl_6/pdata/2/ina/PCtUS9 COMB. seq:PGT-US95-10245-3 +		33.00	87.44	1.9e+03	5428
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-038-682-5 +		33.00	82.51	3.6e+03	9171
/cgnl_6/pdata/2/ina/5A COMB. seq:US-08-302-832-5 +		33.00	82.51	3.6e+03	9171

seq\_name: /cylt\_0/products/2/11a/

seq name: /cqn1 6/ptodata/2/ina/5D COMB.seq:US-08

seq name: /cqn1 6/ptodata/2/ina/5D COMB.seq:US-08-700-636-3

```
seq_documentation_block:
; Sequence 3, Application US/08700636
; Patent No. 5910582
; GENERAL INFORMATION:
; APPLICANT: Eliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,636
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: 08-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9368
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..1553
; US-08-700-636-3

alignment_scores:
Quality: 37.00 Length: 12
Ratio: 3.700 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x US-08-700-636-3
Align seg 1/1 to: US-08-700-636-3 from: 1 to: 1756

1 TyrArgLeuLeuileArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||
732 TACTCGTGATCATCCGGCGGCTGCTGTTGTTCTAC 767

seq_name: /cgn1_6/ptodata/2/ina/5A_COMB.seq:US-08-210-222-8

seq_documentation_block:
; Sequence 8, Application US/08210222
; Patent No. 5599917
; GENERAL INFORMATION:
; APPLICANT: Coppola, George R.
; APPLICANT: Beutel, Bruce A.
; APPLICANT: Bertelsen, Arthur H.
; TITLE OF INVENTION: Inhibition of Interferon- with Oligonucleotides
```

```
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/210,222
; FILING DATE: Unassigned
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Herron, Charles J.
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 23550-114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 BASES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; HYPOTHETICAL: NO
; US-08-210-222-8

alignment_scores:
Quality: 35.00 Length: 11
Ratio: 3.182 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 63.636

alignment_block:
US-08-653-294-17 x US-08-210-222-8
Align seg 1/1 to: US-08-210-222-8 from: 1 to: 98

2 ArgLeuLeuileArgArgIleAlaLeuArgTyr 12
|||||:|||||:|||||:|||||:|||||
30 AGGCUCGUUUUCGCCGGAUCGUCUAGAGUAD 62

seq_name: /cgn1_6/ptodata/2/ina/5C_COMB.seq:US-08-679-405-1

seq_documentation_block:
; Sequence 1, Application US/08679405
; Patent No. 5866393
; GENERAL INFORMATION:
; APPLICANT: Fuglsang, Claus
; APPLICANT: Halkier, Torben
; APPLICANT: Oxenboll, Karen M.
; APPLICANT: Berka, Randy M.
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58663930 No. 5866393dsk of No. 5866393rd America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```



```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/679,405
; FILING DATE: July 9, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA: 08/603,534
; FILING DATE: February 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.210.US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
; US-08-679-405-1
```

```
alignment_scores:
  Quality: 35.00      Length: 12
  Ratio: 3.500      Gaps: 0
  Percent Similarity: 83.333      Percent Identity: 50.000
```

```
alignment_block:
US-08-653-294-17 x US-08-679-405-1 ..
Align seg 1/1 to: US-08-679-405-1 from: 1 to: 2822
1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
||||| :|||||:|||||:|||||:|||||
1377 TACAATCAGATTGTCGTCGATCGCAGTGACTTAC 1412
seq_name: /cgnl_6/ptodata/2/ina/5D_COMB.seq.us-08-842-799-1
```

```
seq_documentation_block:
; Sequence 1, Application US/08842799
; Patent No. 5965418
; GENERAL INFORMATION:
; APPLICANT: Fuglsang, Claus
; APPLICANT: Haikier, Torben
; APPLICANT: Oxenboell, Karen M.
; APPLICANT: Berka, Randy M.
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5965418 of No. 5965418disk of No. 5965418th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/842,799
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,405
; FILING DATE: July 9, 1996
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA: 08/603,534
; FILING DATE: February 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.210.US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
; US-08-842-799-1

alignment_scores:
  Quality: 35.00      Length: 12
  Ratio: 3.500      Gaps: 0
  Percent Similarity: 83.333      Percent Identity: 50.000

alignment_block:
US-08-653-294-17 x US-08-842-799-1 ..
Align seg 1/1 to: US-08-842-799-1 from: 1 to: 2822
1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
||||| :|||||:|||||:|||||:|||||
1377 TACAATCAGATTGTCGTCGATCGCAGTGACTTAC 1412
seq_name: /cgnl_6/ptodata/2/ina/PTUS9_COMB.seq:PCT-US96-11458-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9611458
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11458
; FILING DATE: 9-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,194
```

;; FILING DATE: 14-JUL-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/603,534  
;; FILING DATE: 21-FEB-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Lambiris, Elias J.  
;; REGISTRATION NUMBER: 33,728  
;; REFERENCE/DOCKET NUMBER: 441.204-WO  
;; TELEPHONE: (212) 867-0123  
;; TELEFAX: (212) 878-9655  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2822 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 477..2276  
PCT-US96-11458-1

alignment\_scores:  
Quality: 35.00 Length: 12  
Ratio: 3.500 Gaps: 0  
Percent Similarity: 83.333 Percent Identity: 50.000

## alignment\_block:

US-08-653-294-17 x PCT-US96-11458-1 ..

Align seg 1/1 to: PCT-US96-11458-1 from: 1 to: 2822

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12

||||: ||||:|||||||:||||: |||

1377 TACAATCAGATTGTGCTGCATCGCAGTGACTTAC 1412

seq\_name: /cgnl\_6/ptodata/2/ina/PCTUS9\_COMB.seq:PCT-US96-053320A-897

## seq\_documentation\_block:

;; Sequence 897, Application PC/TUS9605320A

;; GENERAL INFORMATION:

;; APPLICANT: Human Genome Sciences

;; APPLICANT: 9410 Key West Avenue

;; APPLICANT: Rockville, MD 20850

;; APPLICANT: United States of America

;; APPLICANT: Johns Hopkins University

;; APPLICANT: 720 Rutland Avenue

;; APPLICANT: Baltimore, MD 21205

;; APPLICANT: United States of America

;; APPLICANT: Mark D. Adams

;; APPLICANT: Owen White

;; APPLICANT: Hamilton O. Smith

;; APPLICANT: J. Craig Venter

;; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,

;; NUMBER OF SEQUENCES: 48

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox

;; STREET: 1100 New York Avenue, Suite 600

;; CITY: Washington

;; STATE: D.C.

;; COUNTRY: USA

;; ZIP: 20003-3934

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

;; COMPUTER: HP Vectra 486/33

;; OPERATING SYSTEM: MSDOS version 6.2

;; SOFTWARE: ASCII text

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: PCT/US96/053320A

;; FILING DATE: April 22, 1996

;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/476,102  
;; FILING DATE: June 7, 1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/487,429  
;; FILING DATE: June 7, 1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Eric K. Steffe  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1488.014PC01  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 371-2600  
;; TELEFAX: (202) 371-2540  
;; INFORMATION FOR SEQ ID NO: 897:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3096 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
PCT-US96-05320A-897

## alignment\_scores:

Quality: 34.00 Length: 11  
Ratio: 3.400 Gaps: 0  
Percent Similarity: 90.909 Percent Identity: 63.636

## alignment\_block:

US-08-653-294-17 x PCT-US96-05320A-897/rev ..

Align seg 1/1 to reverse of: PCT-US96-05320A-897 from: 1 to: 3096

1 TyrArgLeuLeuIleArgArgIleAlaLeuArg 11

||||: ||||:|||||||:||||: |||

2526 TTCAGATTAAATCAACGTATAGCCTTCGCG 2494

seq\_name: /cgnl\_6/ptodata/2/ina/5C\_COMB.seq:US-08-619-198-2

## seq\_documentation\_block:

;; Sequence 2, Application US/08619198

;; Patent No. 5885831

;; GENERAL INFORMATION:

;; APPLICANT: Young, Michael W.

;; APPLICANT: Sehgal, Amita

;; APPLICANT: Vossball, Leslie B.

;; APPLICANT: Price, Jeffrey L.

;; APPLICANT: Myers, Michael

;; TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED

;; TITLE OF INVENTION: WITH CIRCADIAN RHYTHMS

;; NUMBER OF SEQUENCES: 8

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Klauber &amp; Jackson

;; STREET: 411 Hackensack Avenue

;; CITY: Hackensack

;; STATE: New Jersey

;; COUNTRY: USA

;; ZIP: 07601

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/619,198

;; FILING DATE: 20-MAR-1996

;; CLASSIFICATION: 514

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Jackson Esq., David A.

;; REGISTRATION NUMBER: 26,742

;; REFERENCE/DOCKET NUMBER: 600-1-128A CPI

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 201 487-5800

;; TELEFAX: 201 343-1684

```
;
;
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; DESCRIPTION: tim cdna sequence wherein R at position 1335
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3369
;
US-08-619-198-2

alignment_scores:
  Quality: 34.00 Length: 10
  Ratio: 3.778 Gaps: 0
  Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-17 x US-08-619-198-2/rev ..
Align seg 1/1 to reverse of: US-08-619-198-2 from: 1 to: 3369

      2 ArgLeuLeuIleArgArgIleAlaLeuArg 11
      |||||::: ||:::|||||
1563 AGGCTCTGCTGCTGGCGTGGCCCTGCGG 1534

seq_name: /cgnl_6/ptodata/2/ina/5A_COMB.seq:US-08-121-713D-59

seq_documentation_block:
; Sequence 59, Application US/08121713D
; Patent No. 5639856
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey S.
; APPLICANT: Kolodkin, Alex L.
; APPLICANT: Matthes, David
; APPLICANT: Bentley, David R.
; APPLICANT: O'Connor, Timothy
; TITLE OF INVENTION: The Semaphorin Gene Family
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121.713D
; FILING DATE: 13-SEP-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Osmán, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B94-002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415) 343-4342
; TELEX:
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3560 base pairs
```

```
;
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1953
;
US-08-121-713D-59

alignment_scores:
  Quality: 34.00 Length: 12
  Ratio: 3.778 Gaps: 0
  Percent Similarity: 75.000 Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x US-08-121-713D-59 ..
Align seg 1/1 to: US-08-121-713D-59 from: 1 to: 3560

      1 TyrArgLeuIleArgArgIleAlaLeuArgTyr 12
      |||||::: ||:::|||||
3151 TACACAACTCATATCTACGTCGCGCTGCATAT 3186

seq_name: /cgnl_6/ptodata/2/ina/5B_COMB.seq:US-08-835-268-59

seq_documentation_block:
; Sequence 59, Application US/08835268
; Patent No. 5807826
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey S.
; APPLICANT: Kolodkin, Alex L.
; APPLICANT: Matthes, David
; APPLICANT: Bentley, David R.
; APPLICANT: O'Connor, Timothy
; TITLE OF INVENTION: The Semaphorin Gene Family
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,268
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/121.713
; FILING DATE: 13-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Osmán, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B94-002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415) 343-4342
; TELEX:
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3560 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: 1..1953
US-08-835-268-59

alignment_scores:
  Quality: 34.00      Length: 12
  Ratio: 3.778       Gaps: 0
  Percent Similarity: 75.000  Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x US-08-835-268-59 ..
Align seg 1/1 to: US-08-835-268-59 from: 1 to: 3560

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
||||| :||||| :||||| :|||||
3151 TACACACAACTCATATCTAAGTGGCGCTGGATAT 3186

seq_name: /cgnl_6/ptodata/2/ina/5D_COMB.seq:US-09-060-692-59

seq_documentation_block:
; Sequence 59, Application US/09060692
; Patent No. 5933865
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey S.
; APPLICANT: Kolodkin, Alex L.
; APPLICANT: Matthes, David
; APPLICANT: Bentley, David R.
; APPLICANT: O'Connor, Timothy
; TITLE OF INVENTION: The Semaphorin Gene Family
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,692
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,713
; FILING DATE: 13-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B94-002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415) 343-4342
; TELEX:
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3560 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1953
; US-09-060-692-59

alignment_scores:
  Quality: 34.00      Length: 12
  Ratio: 3.778       Gaps: 0
  Percent Similarity: 75.000  Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x PCT-US94-10151A-59 ..
Align seg 1/1 to: PCT-US94-10151A-59 from: 1 to: 3560

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
||||| :||||| :||||| :|||||
3151 TACACACAACTCATATCTAAGTGGCGCTGGATAT 3186

seq_name: /cgnl_6/ptodata/2/ina/5C_COMB.seq:US-08-619-198-4
```

```
; LOCATION: 1..1953
US-08-835-268-59

alignment_scores:
  Quality: 34.00      Length: 12
  Ratio: 3.778       Gaps: 0
  Percent Similarity: 75.000  Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x US-09-060-692-59 ..
Align seg 1/1 to: US-09-060-692-59 from: 1 to: 3560

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
||||| :||||| :||||| :|||||
3151 TACACACAACTCATATCTAAGTGGCGCTGGATAT 3186

seq_name: /cgnl_6/ptodata/2/ina/PCTUS9_COMB.seq:PCT-US94-10151A-59

seq_documentation_block:
; Sequence 59, Application PC/TUS9410151A
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: The Semaphorin Gene Family
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLERH HOBBACH TEST ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10151A
; FILING DATE: 13-SEP-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-58750-PC/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3560 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1953
; PCT-US94-10151A-59

alignment_scores:
  Quality: 34.00      Length: 12
  Ratio: 3.778       Gaps: 0
  Percent Similarity: 75.000  Percent Identity: 58.333

alignment_block:
US-08-653-294-17 x PCT-US94-10151A-59 ..
Align seg 1/1 to: PCT-US94-10151A-59 from: 1 to: 3560

1 TyrArgLeuLeuIleArgArgIleAlaLeuArgTyr 12
||||| :||||| :||||| :|||||
3151 TACACACAACTCATATCTAAGTGGCGCTGGATAT 3186

seq_name: /cgnl_6/ptodata/2/ina/5C_COMB.seq:US-08-619-198-4
```

```

seq_documentation_block:
; Sequence 4, Application US/08619198
; Patent No. 5885831
; GENERAL INFORMATION:
; APPLICANT: Young, Michael W.
; APPLICANT: Sehgal, Anita
; APPLICANT: Voshall, Leslie B.
; APPLICANT: Price, Jeffrey L.
; APPLICANT: Myers, Michael
; TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED
; WITH CIRCADIAN RHYTHMS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,198
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-128A CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4170 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: tim cDNA sequence wherein R at position 1335
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4170
; US-08-619-198-4

alignment_scores:
Quality: 34.00 Length: 10
Ratio: 3.778 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-17 x US-08-619-198-4/rev ..
Align seg 1/1 to reverse of: US-08-619-198-4 from: 1 to: 4170

2 ArgLeuLeuLeuArgIleAlaLeuArg 11
|||||:|||||:|||||:|||||:
1563 AGGCTTGTGCTGCGGCTTGCCCTGCGG 1534

seq_name: /cgn1_6/ptodata/2/ina/5C_COMB.seq:US-08-619-198-8

seq_documentation_block:
; Sequence 8, Application US/08619198
; Patent No. 5885831
; GENERAL INFORMATION:
; APPLICANT: Young, Michael W.
; APPLICANT: Sehgal, Anita
; APPLICANT: Voshall, Leslie B.
; APPLICANT: Price, Jeffrey L.
; APPLICANT: Myers, Michael
; TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED
; WITH CIRCADIAN RHYTHMS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

```

; GENERAL INFORMATION:
; APPLICANT: Young, Michael W.
; APPLICANT: Sehgal, Anita
; APPLICANT: Voshall, Leslie B.
; APPLICANT: Price, Jeffrey L.
; APPLICANT: Myers, Michael
; TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED
; WITH CIRCADIAN RHYTHMS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,198
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-128A CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5192 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; US-08-619-198-8

alignment_scores:
Quality: 34.00 Length: 10
Ratio: 3.778 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:
US-08-653-294-17 x US-08-619-198-8/rev ..
Align seg 1/1 to reverse of: US-08-619-198-8 from: 1 to: 5192

2 ArgLeuLeuLeuArgIleAlaLeuArg 11
|||||:|||||:|||||:|||||:
1805 AGGCTTGTGCTGCGGCTTGCCCTGCGG 1776

seq_name: /cgn1_6/ptodata/2/ina/5C_COMB.seq:US-08-619-198-1

seq_documentation_block:
; Sequence 1, Application US/08619198
; Patent No. 5885831
; GENERAL INFORMATION:
; APPLICANT: Young, Michael W.
; APPLICANT: Sehgal, Anita
; APPLICANT: Voshall, Leslie B.
; APPLICANT: Price, Jeffrey L.
; APPLICANT: Myers, Michael
; TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED
; WITH CIRCADIAN RHYTHMS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/619,198  
FILING DATE: 20-MAR-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-128A CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5198 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: tim cDNA sequence wherein R at position 1575  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Drosophila melanogaster  
US-08-619-198-1

alignment\_scores:  
Quality: 34.00 Length: 10  
Ratio: 3.778 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 70.000

alignment\_block:  
US-08-653-294-17 x US-08-619-198-1/rev ..

Align seg 1/1 to reverse of: US-08-619-198-1 from: 1 to: 5198

2 ArgLeuLeuIleArgArgIleAlaLeuArg 11  
|||||  
1803 AGGCTTCTGCTGCGGCTTGCCTGCGG 1774

OM of: US-08-653-294-17 to: EST:\* out\_format : pfs  
Date: Feb 8, 2000 4:03 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame\_param\_p2n.model -DEV=xlip  
-Q=/cgnl1/USP10.spool/US08653294/runat\_04022000\_160700/app\_query.fasta.1  
-DB=EST -PMIT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MAPFIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-653-294-17

Query length: 12

Database: EST:\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 8553.360000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_gss6:A0851612	-	43.00	141.90	38.74	724	AQ851612 Cpg1352B Cptowagdnal
gb_est24:A1210579	-	41.00	140.22	48.07	377	A1210579 J9c12a1 rl aspergillus
gb_gss11:A0429279	+	41.00	137.54	67.75	513	AQ429279 CIBI-EL-2565K5-TR CIT
gb_est26:A0005267	+	41.00	133.58	112.71	810	AU005267 AU005267 Bombyx mori P
gb_est27:A1426843	+	40.00	139.41	53.31	271	A1426843 mm95b08.x1 Stratagene
gb_est36:A182651	+	40.00	138.53	59.71	300	AV182651 AV182651 Yuji Kohara
gb_gss8:AQ015503	+	40.00	135.06	93.11	447	AQ015503 CIT-HSP-2309J17-TR CIT
gb_est39:A1271330	+	39.00	134.57	99.23	310	AW127330 M110501 DSIL Medicago
gb_gss6:AQ883371	+	39.00	129.84	181.89	534	AQ883371 HS_5471_B2_C05_T7A RPC
gb_gss11:AQ433893	+	39.00	129.45	191.40	559	AQ433893 HS_5036_A1_D05_T7 RPC
gb_gss11:AG000734	+	39.00	127.92	232.65	666	AG000734 Homo sapiens genomic D
gb_gss11:AQ0038101	+	38.00	134.66	98.11	201	AQ238101 RPII1-73J18.TK RPII-1
gb_gss11:CN500790	+	38.00	129.62	187.24	359	AL083934 Arabidopsis thaliana
gb_gss9:AQ166595	+	38.00	128.99	202.99	386	HS_3146_A1_B09_MR CIT
gb_gss9:AQ137850	+	38.00	127.45	247.11	451	HS_3058_B2_C09_MF CIT
gb_gss10:AQ242894	+	38.00	127.20	255.19	474	AQ242894 HS_2061_A1_C04_MR CIT
gb_gss8:AQ055939	+	38.00	126.35	284.76	523	AQ055939 CIT-HSP-2340F9-TF CIT
gb_est37:A1949481	+	38.00	125.96	299.36	547	A1949481 w10g04.x1 NCI-CGAP-P
gb_est31:A1678132	+	38.00	125.48	318.32	578	A1678132 w23f04.x1 NCI-CGAP-P
gb_est22:A1058050	+	38.00	125.24	328.16	594	A1058050 SWAMC12C11sk Brugia
gb_est23:A1096218	+	38.00	125.20	330.00	597	AT096218 SWAMC26G06SK Brugia
gb_gss6:AQ835306	+	38.00	125.08	334.94	605	AQ835306 HS_4815_A1_D12_T7A CIT
gb_gss11:CN500809	+	38.00	123.50	410.39	726	AL051513 Drosophila melanogaste
gb_gss8:AQ060698	+	37.50	125.55	315.44	464	AQ060698 CIT-HSP-2346D8-TR CIT
gb_est31:F29413	-	37.00	128.16	225.62	278	F29413 HSP19257 HM3 Homo sapie
gb_est31:F24270	-	37.00	127.86	234.69	288	F24270 HSP10415 HM3 Homo sapie
gb_est16:C26415	+	37.00	127.39	249.26	304	C26415 C26415 Rice callus cDNA
gb_est18:AA523786	+	37.00	127.27	252.92	308	AA523786 n168f08.s1 NCI-CGAP-P
gb_est19:AA748739	+	37.00	127.25	253.83	309	AA748739 ny06e12.s1 NCI-CGAP-G
gb_est44:AW197413	+	37.00	126.86	266.68	323	AW197413 x42g08.x1 NCI-CGAP-G
gb_est10:AA182557	+	37.00	126.83	267.60	324	AA182557 x256g03.s1 Stratagene
gb_est24:A1202172	+	37.00	126.75	270.36	327	AT202172 Q569b08.x1 NCI-CGAP-P
gb_est33:AV090309	+	37.00	126.70	272.21	329	AV090309 AV090309 Mus musculus
gb_gss3:B60584	+	37.00	126.14	292.37	351	B60584 CIT-HSP-2013F20-TRC CIT
gb_gss8:AQ075747	+	37.00	125.92	300.94	360	AQ075747 CIT-HSP-2354G2-TR CIT
gb_gss6:AQ826246	+	37.00	125.87	302.80	362	AQ826246 HS_5555_A2_E12.SP6 RF
gb_gss3:B31543	+	37.00	125.54	315.88	376	B31543 HS-1011-A2-A06-MF abi CI
gb_est11:AA265909	+	37.00	125.00	338.43	400	AA265909 ms72c02.r1 Soares mous
gb_est16:AA572328	-	37.00	124.64	354.49	417	AA572328 v188a05.r1 Stratagene
gb_est22:A1050480	-	37.00	124.52	360.18	423	AW050480 uc86g01.y1 Sugano mous
gb_est38:AW091827	-	37.00	124.50	361.13	424	AW091827 EST285007 tomato mixed
gb_est24:AU031224	+	37.00	124.29	370.63	434	AU031224 AU031224 Rice cDNA fr

gb\_gss3:B74688 + 37.00 124.17 376.35 440 ! B74688 CIT-HSP-2034E11-TF CI  
gb\_gss5:AQ815539 - 37.00 124.04 383.03 447 ! AQ815539 HS\_5348\_A2\_A09.SP6E  
gb\_est15:AA478141 - 37.00 124.02 383.98 448 ! AA478141 ZU42E09.r1 Soares o

seq\_name: gb\_gss6:AQ851612

seq\_documentation\_block:

LOCUS AQ851612 724 bp DNA GSS 18-OCT-1999  
DEFINITION Cpg1352B Cptowagdnal Cryptosporidium parvum genomic similar to SKB1  
homologous (negative regulator of mitosis) (regulator of Shk1, a  
p21(Cdc42/Rac)-activated kinase (PAK)), genomic survey sequence.  
ACCESSION AQ851612  
VERSION AQ851612  
KEYWORDS GSS.  
SOURCE Cryptosporidium parvum.  
ORGANISM Cryptosporidium parvum  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
Cryptosporidiidae; Cryptosporidium.  
REFERENCE 1 (bases 1 to 724)  
AUTHORS Strong,W.B. and Nelson,R.G.  
TITLE Cryptosporidium parvum GSS Project  
JOURNAL Unpublished (1997)  
COMMENT On Sep 10, 1998 this sequence version replaced gi:3553959.  
Contact: Nelson, R. G.  
Depts. of Medicine & Pharmaceutical Chemistry  
San Francisco General Hospital-University of California, San  
Francisco  
Box 0811, San Francisco, CA 94143-0811, USA  
Tel: 415 206 8846  
Fax: 415 206 3353  
Email: malari@itsa.ucsf.edu

For Annotation Data see <http://medsfgh.ucsf.edu/id/CpTags/home.html>  
Seq primer: T3  
Class: Shotgun.  
Location/Qualifiers  
1..724  
/organism="Cryptosporidium parvum"  
/strain="IOWA"  
/db\_xref="taxon:5807"  
/clone\_lib="CpIOWAGNAL"  
/lab\_host="E. coli XL2 Blue MRF"  
/note="Vector: pBluescript II (SK-); Site1: EcoRV; C.  
parvum (IOWA isolate) genomic DNA was hydrodynamically  
sheared to produce fragments having a tight size  
distribution between 2-4 kb by Dr. Yvonne Thorstenson of  
the Stanford DNA Sequencing and Technology Center  
([http://sequence-  
www.stanford.edu/group/techdev/shear.htm](http://sequence-www.stanford.edu/group/techdev/shear.htm)). The randomly  
sheared gDNA was chromatographed on Sephacryl S-400 to  
remove any small fragments and DNA eluting in the void  
volume was subcloned into an EcoR V-digested, alkaline  
phosphate-treated pBluescript II (SK-) vector and  
transformed into E. coli strain XL2 Blue MRF.  
Recombinant clones from the first plating of the library  
were selected for sequence analysis using T3 and T7  
primers."

BASE COUNT 264 a 93 c 121 g 244 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 43.00 Length: 11  
Ratio: 4.300 Gaps: 0  
Percent Similarity: 90.909 Percent Identity: 90.909

alignment\_block:  
US-08-653-294-17 x AQ851612/rev ..  
Align seg 1/1 to reverse of: AQ851612 from: 1 to: 724

1 TyrArgLeuLeuLeuArgArgLeuLeuArg 11  
||||||| ||||||| ||||||| |||||||  
552 TATAGCTCTAATAAGACGTATAGCTTGAGG 520

```

seq_name: gb_est24:A1210579
seq_documentation_block: 377 bp mRNA EST 19-OCT-1998
LOCUS A1210579 j9c12a1.r1 Aspergillus nidulans 24hr asexual developmental and
DEFINITION vegetative cDNA lambda zap library Emericella nidulans cDNA clone
j9c12a1 5', mRNA sequence.
ACCESSION A1210579 GI:3772521
VERSION A1210579.1
KEYWORDS EST.
SOURCE Emericella nidulans.
ORGANISM Emericella nidulans
Eukaryota: Fungi; Ascomycota; Eurotiales; Trichocomaceae;
Emericella
REFERENCE 1 (bases 1 to 377)
AUTHORS Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
Prade, R. and Roe, B.
TITLE An Aspergillus nidulans EST Database
JOURNAL Unpublished (1998)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2285706.
Other_ESRs: j9c12a1.f1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: T3.
FEATURES
SOURCE
1..377
/organism="Emericella nidulans"
/strain="FGSC A26"
/db_xref="taxon:5072"
/clone="j9c12a1"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/tissue_type="vegetative mycelia, asexual structures"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 74 a 127 c 103 g 73 t
ORIGIN

alignment_scores:
Quality: 41.00 Length: 11
Ratio: 3.727 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 72.727

alignment_block:
US-08-653-294-17 x A1210579/rev ..
Align seg 1/1 to reverse of: A1210579 from: 1 to: 377

1 TyArgLeuLeuIleArgArgIleAlaLeuArg 11
:::|||||:|||||:|||||:|||||:|||||:
213 CATCGCTGTGCTGCGCGGATGATCCCTCGCG 181

seq_name: gb_gss13:AQ429279
seq_documentation_block: 513 bp DNA GSS 24-MAR-1999
LOCUS AQ429279 CITBI-EI-2565K5.TR CITBI-EI Homo sapiens genomic clone 2565K5,
DEFINITION genomic survey sequence.
ACCESSION AQ429279
VERSION AQ429279.1 GI:4497045
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 513)
AUTHORS Zhao, S., Adams, M.D., Niernan, W., Malek, J., Shizuya, H., Simon, M. and
Venter, J.C.
TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1997)
Other_GSSs: CITBI-EI-2565K5.TF
Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbettigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
SOURCE
1..513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2565K5"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 162 a 128 c 101 g 122 t
ORIGIN

alignment_scores:
Quality: 41.00 Length: 11
Ratio: 4.100 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 81.818

alignment_block:
US-08-653-294-17 x AQ429279 ..
Align seg 1/1 to: AQ429279 from: 1 to: 513

1 TyArgLeuLeuIleArgArgIleAlaLeuArg 11
|||||:|||||:|||||:|||||:|||||:
7 TACCGAGCTTTAATTCACCGCATAGCTCTCAGA 39

seq_name: gb_est26:AU005267
seq_documentation_block: 810 bp mRNA EST 19-JAN-1999
LOCUS AU005267 Bombyx mori p50(Daizo) Bombyx mori cDNA clone ws30474,
DEFINITION mRNA sequence.
ACCESSION AU005267
VERSION AU005267.1 GI:4162638
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 810)
AUTHORS Mita, K., Morinoy, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150717.
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp

```



Align seq 1/1 to: AV182651 from: 1 to: 300

and





```

MEDLINE
COMMENT
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3146 row: C column: 17
Class: BAC ends
High quality sequence stop: 386.
Location/Qualifiers
1..386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3146 Col-17 Row-C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 98 a 65 C 75 g 148 t
ORIGIN

alignment_scores:
Quality: 38.00 Length: 11
Ratio: 3.800 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 54.545

alignment_block:
US-08-653-294-17 x AQ166595 ..
Align seg 1/1 to: AQ166595 from: 1 to: 386

2 ArgLeuLeuLeuArgGileAlaLeuArgTyr 12
142 CGATTAAATGCTGAGAGATTGGCTATAGATAT 174

seq_name: gb_gss9:AQ137850

seq_documentation_block:
LOCUS AQ137850 461 bp DNA GSS 24-SEP-1998
DEFINITION HS_3058_P2_C09_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3058 Col=18 Row=F, genomic survey
sequence.
ACCESSION AQ137850
VERSION AQ137850.1 GI:3528503
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 461)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999).
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3058 row: F column: 18
Class: BAC ends
High quality sequence stop: 461.
Location/Qualifiers

```

```

source
1..461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3058 Col-18 Row-F"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 132 a 125 C 102 g 102 t
ORIGIN

alignment_scores:
Quality: 38.00 Length: 12
Ratio: 4.222 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 66.667

alignment_block:
US-08-653-294-17 x AQ137850 ..
Align seg 1/1 to: AQ137850 from: 1 to: 461

1 TyrArgLeuLeuLeuArgGileAlaLeuArgTyr 12
:::||||| |||||||||
18 CATCGACTCACTAAGGCGCAATTCGGCCAGGTAC 53

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:29:40 : Search time 122.56 Seconds  
(without alignments)  
2.319 Million cell updates/sec

Title: US-08-653-294-18

Perfect score: 58

Sequence: 1 YRLAIRILLRY 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_36;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	91.4	12	1 R95429	HLA-B2702 84-79-84
2	53	91.4	12	1 W33798	Peptide B2702.84-7
3	53	91.4	12	1 W33799	Immunomodulating d
4	39	67.2	20	1 R92907	HLA-B2702 CTL modu
5	39	67.2	20	1 R95428	HLA-B2702 84-75-84
6	39	67.2	20	1 W33778	Immunomodulating d
7	34	58.6	20	1 R92909	HLA-B2702 CTL modu
8	34	58.6	20	1 R92908	HLA-B2702 CTL modu
9	34	58.6	20	1 W33791	Peptide B2702.84-7
10	34	58.6	20	1 W33792	Peptide B2702.84-7
11	32	55.2	20	1 R95430	HLA-B2702 84-75T/7
12	32	55.2	487	1 R14149	Human alpha 2 beta
13	32	55.2	487	1 W11804	Human alpha-2b adr
14	32	55.2	504	1 W44156	Human neuronal nic
15	31.5	54.3	803	1 W55668	H. pylori ORF hp4e
16	31.5	54.3	804	1 W55368	H. pylori ORF 09ap
17	31.5	54.3	1123	1 Y07119	Lung cancer associ
18	31.5	54.3	1177	1 Y07120	Lung cancer associ
19	31	53.4	216	1 R52029	Protein with Oreta
20	31	53.4	350	1 W62821	Mus musculus SOC7
21	31	53.4	469	1 W48382	Homo sapiens don-1
22	31	53.4	546	1 W92336	M. rosa truncated
23	31	53.4	564	1 W92335	M. rosa fourth adh
24	31	53.4	605	1 W48379	Mus musculus don-1
25	31	53.4	647	1 W48383	Homo sapiens don-1
26	31	53.4	860	1 W63700	Receptor type tyro
27	31	53.4	4473	1 R97244	Virulence gene clu
28	30	51.7	18	1 R71429	Human MHC 1 alpha
29	30	51.7	119	1 R74034	Bombesin-related p
30	30	51.7	362	1 R48695	G-protein coupled
31	30	51.7	362	1 W02667	G-protein coupled
32	30	51.7	402	1 R06495	Beta 3 adrenergic
33	30	51.7	405	1 R62515	Bovine beta3 adren
34	30	51.7	405	1 W44933	Canine beta-3 adre

35 30 51.7 407 1 W53847 Human adrenaline b  
36 30 51.7 408 1 R54991 Human beta-3 adren  
37 30 51.7 428 1 R39264 Murine somatostati  
38 30 51.7 627 1 W44152 Human neuronal nic  
39 30 51.7 686 1 R25591 RING11 antigenic p  
40 29.5 50.9 20 1 R92911 HLA-B2702 CTL modu  
41 29.5 50.9 20 1 W33779 Immunomodulating d  
42 29 50.0 6 1 W47263 Immunomodulatory p  
43 29 50.0 6 1 W47262 Immunomodulatory p  
44 29 50.0 6 1 W33782 Peptide #3 used in  
45 29 50.0 6 1 W33781 Peptide #2 used in

## ALIGNMENTS

RESULT 1

R95429 ID R95429 standard; peptide; 12 AA.  
AC R95429;  
DT 12-NOV-1996 (first entry)  
DE HLA-B2702 84-79-84 palindrome.  
KW HLA; p74; alpha-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN W0513288-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PI (STRD ) UNIV LELAND STANFORD JUNIOR.  
PA Clayberger C. Krensky AM;  
WP1: 95-194027/25  
DT Compens. comprising lymphoid surface membrane proteins - which may  
PT Inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12; 29pp; English.  
CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. These sequences can be used to isolate  
CC HLA-B2702 84-79-84 palindrome. The protein p74 is a T-cell surface membrane  
CC protein associated with T-cell activation in mammalian T-cells, and is  
CC also immunologically cross reactive with the heat shock protein Hsc70.  
CC p74 is found in a limited number of cell types, but is particularly  
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
CC cell with an amphoteric detergent, and then passed through an affinity  
CC column containing a covalently bound HLA-B2702 palindromic peptide.  
CC Compositions comprising the extracellular fragment of p74 combined with  
CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits  
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate  
CC compounds can be screened for their effect on the cytolytic activity of  
CC T-cells, by combining them with the extracellular portion of p74 and  
CC determining the amount of binding between the candidate compound and p74.  
CC Modulation of CTL activity can be inhibited in a cellular composition  
CC containing T-cells and antigen presenting cells (APCs), by adding to the  
CC mix the extracellular portion of p74, in an amount sufficient to compete  
CC with p74 for the binding of the p74 ligand.  
SQ Sequence 12 AA;

Query Match 91.4%; Score 53; DB 1; Length 12;  
Best Local Similarity 91.7%; Pred. No. 0.00042;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLAIRILLRY 12

DB 1 YRLAIRILLRY 12

RESULT 2

W33798

ID W33798 standard; peptide; 12 AA.

AC W33798;

DT 19-JUN-1998 (first entry)  
 DE Peptide B2702.84-79/79-84 tested for immunomodulating activity.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997; U08689.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Example 1; Page 19; 41pp; English.  
 CC Peptides W3784-98 and W3778-9 were assayed for their immunomodulating  
 CC activity. A peptide-type compound or variant is claimed which has  
 CC immunomodulating activity, including the N-terminal acylated and/or  
 CC C-terminal amidated or esterified forms of up to 60 amino acids, where  
 CC the peptide-type compound comprises the formula: A-B, where A, B =  
 CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =  
 CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
 CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino  
 CC acid. The sequence in the brackets may optionally be absent or truncated  
 CC at any peptide type bond within the brackets. The compounds comprise  
 CC amino acid sequences related to a Class I HLA-B alpha domain (positions  
 CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from  
 CC undesirably attacking cells in a host or in vitro. They can also be  
 CC used in combination with antigenic peptides or proteins of interest to  
 CC activate CTLs. They can also inhibit the proliferation of T cells in  
 CC response to anti-CD3. The peptide can be used for preventing rejection  
 CC of transplants or for treating autoimmune diseases, e.g. diabetes,  
 CC rheumatoid arthritis and lupus erythematosus. The products can also be  
 CC used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 91.4%; Score 53; DB 1; Length 12;  
 Best Local Similarity 91.7%; Pred. No. 0.00042;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLAIRRILLRY 12  
 |||||  
 DB 1 YRLAIRRIALRY 12

RESULT 3  
 W3799  
 ID W3799 standard; peptide; 12 AA.  
 AC W3799;  
 DT 19-JUN-1998 (first entry)  
 DE Immunomodulating dimer peptide #3.  
 KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
 KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
 KW rejection.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9744351-A1.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08689.  
 PR 24-MAY-1996; US-653294.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Beulow R, Clayberger C, Krensky AM;  
 DR WPI: 98-086530/08.  
 PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
 alpha-1 domain, used for preventing rejection of transplants or  
 PT treating autoimmune diseases  
 PS Claim 17; Page 35; 41pp; English.  
 CC This sequence represents a specifically claimed immunomodulating  
 CC dimer peptide of the invention. A peptide-type compound or variant is

CC claimed which has immunomodulating activity, including the N-terminal  
 CC acylated and/or C-terminal amidated or esterified forms of up to 60  
 CC amino acids, where the peptide-type compound comprises the formula: A-B,  
 CC where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or  
 CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
 CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa  
 CC represents amino acid. The sequence in the brackets may optionally be  
 CC absent or truncated at any peptide type bond within the brackets. The  
 CC compounds comprise amino acid sequences related to a Class I HLA-B  
 CC alpha domain (positions 79-84). They can be used to inhibit cytotoxic  
 CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in  
 CC vitro. They can also be used in combination with antigenic peptides or  
 CC proteins of interest to activate CTLs. They can also inhibit the  
 CC proliferation of T cells in response to anti-CD3. The peptide can be  
 CC used for preventing rejection of transplants or for treating autoimmune  
 CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.  
 CC The products can also be used for detection and diagnosis.  
 SQ Sequence 12 AA;

Query Match 91.4%; Score 53; DB 1; Length 12;  
 Best Local Similarity 91.7%; Pred. No. 0.00042;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLAIRRILLRY 12  
 |||||  
 DB 1 YRLAIRRIALRY 12

## RESULT 4

R92907  
 ID R92907 standard; peptide; 20 AA.  
 AC R92907;  
 DT 16-MAY-1996 (first entry)  
 DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).  
 KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
 KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
 KW class I MHC; HLA-B2702.  
 OS Synthetic.  
 PN WO9526979-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1995; U04349.  
 PR 05-APR-1994; US-222851.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Clayberger C, Krensky AM, Parham P;  
 DR WPI: 95-358582/46.  
 PT Extension of acceptance period of transplants from MHC unmatched  
 PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
 PT host  
 PS Example 15; Page 36; 80pp; English.  
 CC R83051-R83085, R83090-R83096 and R92907-R92914 represent fragments of  
 CC class I major histocompatibility complex (MHC) antigens. This sequence  
 CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
 CC I MHC HLA-B2702. These sequences can be used to extend the period of  
 CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
 CC The peptides are administered to a patient in conjunction with a  
 CC subtherapeutic amount of an immunosuppressant. This is administered to  
 CC the patient for a limited period of time (compared to the lifetime  
 CC administration for current treatments). The peptides particularly  
 CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
 CC of the patient.  
 SQ Sequence 20 AA;

Query Match 67.2%; Score 39; DB 1; Length 20;  
 Best Local Similarity 55.0%; Pred. No. 0.28;  
 Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLRY 12  
 |||||  
 DB 1 YRLAIRLNRRNLRIALRY 20



RESULT 5  
R95428 ID R95428 standard; peptide; 20 AA.  
AC R95428 DT 12-NOV-1996 (first entry)  
DE HLA-B\*2702 84-75-84 palindromic.  
KW HLA; p74; alpha-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN WO9513288-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM;  
DR WPI; 95-194027/25.  
PT Compsns. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12: 29pp; English.  
CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. This sequence represents the  
CC HLA-B\*2702 84-75-84 palindromic. These sequences can be used to isolate  
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane  
CC protein associated with T-cell activation in mammalian T-cells, and is  
CC also immunologically cross reactive with the heat shock protein Hsc70.  
CC p74 is found in a limited number of cell types, but is particularly  
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable  
CC cell with an amphoteric detergent, and then passed through an affinity  
CC column containing a covalently bound HLA-B\*2702 palindromic peptide.  
CC Compositions comprising the extracellular fragment of p74 combined with  
CC HLA-B\*2702 60-84 (see R95416), induces calcium influx, and inhibits  
CC cytotoxic T lymphocyte (CTL) differentiation or cytotoxicity. Candidate  
CC compounds can be screened for their effect on the cytolytic activity of  
CC T-cells, by combining them with the extracellular portion of p74 and  
CC determining the amount of binding between the candidate compound and p74.  
CC Modulation of CTL activity can be inhibited in a cellular composition  
CC containing T-cells and antigen presenting cells (APCs), by adding to the  
CC mix the extracellular portion of p74, in an amount sufficient to compete  
CC with p74 for the binding of the p74 ligand.  
SQ Sequence 20 AA;

Query Match 67.28; Score 39; DB 1; Length 20;  
Best Local Similarity 55.08; Pred. No. 0.28;  
Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLY 12  
| | | | |  
Db 1 YRLAIRLERERLERIALRY 20

RESULT 6  
W33778 ID W33778 standard; peptide; 20 AA.  
AC W33778 DT 19-JUN-1998 (first entry)  
DE Immunomodulating dimer peptide #1.  
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;  
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;  
KW rejection.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9744351-A1.  
PD 27-NOV-1997.  
PF 22-MAY-1997; U08689.  
PR 24-MAY-1996; US-653294.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Beulow R, Clayberger C, Krensky AM;  
DR WPI; 98-086530/08.  
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B  
PT alpha-1 domain, used for preventing rejection of transplants or

PT treating autoimmune diseases  
PS Claim 16; Page 35; 41pp; English.  
CC This sequence represents a specifically claimed immunomodulating  
CC dimer peptide of the invention. A peptide-type compound or variant is  
CC claimed which has immunomodulating activity, including the N-terminal  
CC acylated and/or C-terminal amidated or esterified forms of up to 60  
CC amino acids, where the peptide-type compound comprises the formula: A-B,  
CC where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or  
CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a  
CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa  
CC represents amino acid. The sequence in the brackets may optionally be  
CC absent or truncated at any peptide type bond within the brackets. The  
CC compounds comprise amino acid sequences related to a Class I HLA-B  
CC alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic  
CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in  
CC vitro. They can also be used in combination with antigenic peptides or  
CC proteins of interest to activate CTLs. They can also inhibit the  
CC proliferation of T cells in response to anti-CD3. The peptide can be  
CC used for preventing rejection of transplants or for treating autoimmune  
CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosus.  
CC The products can also be used for detection and diagnosis.  
SQ Sequence 20 AA;

Query Match 67.28; Score 39; DB 1; Length 20;  
Best Local Similarity 55.08; Pred. No. 0.28;  
Matches 11; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLY 12  
| | | | |  
Db 1 YRLAIRLERERLERIALRY 20

RESULT 7  
R92909 ID R92909 standard; peptide; 20 AA.  
AC R92909 DT 16-MAY-1996 (first entry)  
DE HLA-B\*2702 CTL modulating peptide (B2702.84-75/75-84(T)).  
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;  
KW immunosuppressant; graft versus host disorder; transplantation; therapy;  
KW Class I MHC; HLA-B\*2702.  
OS Synthetic.  
PN WO9526979-A1.  
PD 12-OCT-1995.  
PF 05-APR-1995; U04349.  
PR 05-APR-1994; US-222851.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM, Parham P;  
DR WPI; 95-358582/46.  
PT Extension of acceptance period of transplants from MHC unmatched  
PT donor hosts - using Class I B75-84 MHC antigen of the recipient  
PT host  
PS Example 15; Page 36; 80pp; English.  
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of  
CC Class I major histocompatibility complex (MHC) antigens. This sequence  
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class  
CC I MHC HLA-B\*2702. These sequences can be used to extend the period of  
CC acceptance by a recipient of a transplant from an MHC unmatched donor.  
CC The peptides are administered to a patient in conjunction with a  
CC subtherapeutic amount of an immunosuppressant. This is administered to  
CC the patient for a limited period of time (compared to the lifetime  
CC administration for current treatments). The peptides particularly  
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)  
CC of the patient.  
SQ Sequence 20 AA;

Query Match 58.68; Score 34; DB 1; Length 20;  
Best Local Similarity 50.08; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLY 12

```

DB 1 YRLAIRLNRERLRLALRY 20
|||||
1 YRLAIRLNRERLRLALRY 20

RESULT 8
ID R92908 standard; peptide; 20 AA.
AC R92908;
DT 16-MAY-1996 (first entry)
DE HLA-B*2702 CTL modulating peptide (B2702.84-75(T)/75-84).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLA-B*2702.
OS Synthetic.
PN W09526979-AL.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.
PR 05-APR-1994; US-222851.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI: 95-358582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B*5-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC I MHC HLA-B*2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLY 12
|||||
DB 1 YRLAIRLNRERLRLALRY 20

RESULT 10
ID W33792 standard; peptide; 20 AA.
AC W33792;
DT 19-JUN-1998 (first entry)
DE Peptide B2702.84-75/75-84T tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN W09744351-AL.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI: 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise
CC amino acid sequences related to a Class I HLA-B alpha domain (positions
CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC undesirably attacking cells in a host or in vitro. They can also be
CC used in combination with antigenic peptides or proteins of interest to
CC activate CTLs. They can also inhibit the proliferation of T cells in
CC response to anti-CD3. The peptide can be used for preventing rejection
CC of transplants or for treating autoimmune diseases, e.g. diabetes,
CC rheumatoid arthritis and lupus erythematosus. The products can also be
CC used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLY 12
|||||
DB 1 YRLAIRLNRERLRLALRY 20

RESULT 10
ID W33792 standard; peptide; 20 AA.
AC W33792;
DT 19-JUN-1998 (first entry)
DE Peptide B2702.84-75/75-84T tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN W09744351-AL.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI: 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula: A-B, where A, B =
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = a hydrophobic or
CC small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC acid. The sequence in the brackets may optionally be absent or truncated
CC at any peptide type bond within the brackets. The compounds comprise
CC amino acid sequences related to a Class I HLA-B alpha domain (positions
CC 79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC undesirably attacking cells in a host or in vitro. They can also be
CC used in combination with antigenic peptides or proteins of interest to
CC activate CTLs. They can also inhibit the proliferation of T cells in
CC response to anti-CD3. The peptide can be used for preventing rejection
CC of transplants or for treating autoimmune diseases, e.g. diabetes,
CC rheumatoid arthritis and lupus erythematosus. The products can also be
CC used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 58.6%; Score 34; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLY 12
|||||
DB 1 YRLAIRLNRERLRLALRY 20

RESULT 10
ID W33791 standard; peptide; 20 AA.
AC W33791;
DT 19-JUN-1998 (first entry)
DE Peptide B2702.84-75/75-84 tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN W09744351-AL.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI: 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or

```

Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLR 12  
| | | | | | | | | |  
Db 1 YRLAIRLNRETNRLRALR 20

## RESULT 11

ID R95430 standard; peptide; 20 AA.  
AC R95430;  
DT 12-NOV-1996 (first entry)  
DE HLA-B2702 84-75T75-84T palindrome.  
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;  
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;  
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;  
KW cytolysis; antigen presenting cell.  
OS Synthetic.  
PN W09513288-A1.  
PD 18-MAY-1995.  
PF 10-NOV-1994; U12985.  
PR 10-NOV-1993; US-150493.  
PA (STRD) UNIV LELAND STANFORD JUNIOR.  
PI Clayberger C, Krensky AM;  
DR WPI; 95-194027/25.  
PT Compsns. comprising lymphoid surface membrane proteins - which may  
PT inhibit cytolytic activity and differentiation of CTLs.  
PS Example; Page 12; 29pp; English.  
CC R95413, and R95415-R95431 represent palindromes and fragments of  
CC human-leucocyte-associated antigens. This sequence represents the  
CC HLA-B2702 84-75T75-84T palindrome. These sequences can be used to  
CC isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface  
CC membrane protein associated with T-cell activation in mammalian T-cells,  
CC and is also immunologically cross reactive with the heat shock protein  
CC Hsc70. p74 is found in a limited number of cell types, but is  
CC particularly expressed on B and T cells. p74 can be isolated by lysis of  
CC a suitable cell with an amphoteric detergent, and then passed through an  
CC affinity column containing a covalently bound HLA-B2702 palindromic  
CC peptide. Compositions comprising the extracellular fragment of p74  
CC combined with HLA-B2702.60-84 (see R95416), induces calcium influx, and  
CC inhibits cytotoxic T lymphocyte (CTL) differentiation or cytotoxicity.  
CC candidate compounds can be screened for their effect on the cytolytic  
CC activity of T-cells, by combining them with the extracellular portion of  
CC p74 and determining the amount of binding between the candidate compound  
CC and p74. Modulation of CTL activity can be inhibited in a cellular  
CC composition containing T-cells and antigen presenting cells (APCs), by  
CC adding to the mix the extracellular portion of p74, in an amount  
CC sufficient to compete with p74 for the binding of the p74 ligand.  
SQ Sequence 20 AA:

Query Match 55.2%; Score 32; DB 1; Length 20;  
Best Local Similarity 52.6%; Pred. No. 5.7;  
Matches 10; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 YRLAIR-----RILLR 11  
| | | | | | | | | |  
Db 1 YRLAIRLNRETNRLRALR 19

## RESULT 12

ID R14149 standard; Protein; 487 AA.  
AC R14149;  
DT 06-JAN-1992 (first entry)  
DE Human alpha 2 beta adrenergic receptor.  
KW Neurotransmission; adrenaline; epinephrine; NGC-alpha2beta.  
OS Homo sapiens.  
PN U53053337-A.  
PD 01-OCT-1991.  
PF 30-OCT-1989; 428856.  
PR 30-OCT-1989; US-428856.

PA (NEUR-) NEUROGENETIC CORP.  
PI Weinshank RL, Hartig PR;  
DR WPI; 91-310087/42.  
DR N-PSDB; Q14151.

PT Isolated DNA encoding human adrenergic receptor - for detecting  
PT nucleic acids encoding alpha, 2-beta adrenergic receptor, for  
PT screening drugs.  
PS Disclosure; Fig 2; 15pp; English.  
CC Clone NGC-alpha2beta was isolated from a human spleen genomic  
CC library by screening with a fragment of the human 5-HT1A receptor  
CC gene. The gene was used to express recombinant receptor protein  
CC which can be used to produce antibodies for inhibition of receptor  
CC function.  
SQ Sequence 487 AA:

Query Match 55.2%; Score 32; DB 1; Length 487;  
Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLAIRRILLR 11  
: | | | | | | | | | |  
Db 470 FRAPRRILCR 480

## RESULT 13

ID W11804 standard; Protein; 487 AA.  
AC W11804;  
DT 06-MAY-1997 (first entry)  
DE Human alpha-2b adrenergic receptor.  
KW Alpha-2b adrenergic receptor; adrenoceptor; adrenaline;  
KW epinephrine; signal transduction; neurotransmitter; ligand.  
OS Homo sapiens.  
PN U5595880-A.  
PD 21-JAN-1997.  
PF 30-OCT-1989; 428856.  
PR 30-OCT-1989; US-428856.  
PR 30-MAY-1991; US-707604.  
PR 22-OCT-1992; US-965040.  
PA (SYNA-) SYNAPTIC PHARM CORP.  
PI Hartig PR, Weinshank RL;  
DR WPI; 97-107576/10.  
DR N-PSDB; T59499.  
PT Assay for alpha-2b adrenergic receptor ligands - using membranes of  
PT cells expressing recombinant receptor  
PS Disclosure; Fig 2A-E; 16pp; English.  
CC Human alpha-2b adrenergic receptor (W11804) is a member of the  
CC rhodopsin-like signal transducer family. Its amino acid sequence  
CC was deduced from a genomic DNA clone (T59499) obtd. from a human  
CC spleen DNA library. Vectors have been adapted to allow prodn. of  
CC alpha-2b adrenoceptor in bacterial, yeast or mammalian cells;  
CC transfected Itk- cells, designated L-NGC-alpha-2B, are deposited as  
CC ATCC CRL 10275. Membranes of transfected mammalian cells can used  
CC in novel methods to identify drugs which specifically interact  
CC with, and bind to, the alpha-2b adrenergic receptor.  
SQ Sequence 487 AA:

Query Match 55.2%; Score 32; DB 1; Length 487;  
Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRLAIRRILLR 11  
: | | | | | | | | | |  
Db 470 FRAPRRILCR 480

## RESULT 14

ID W44156 standard; Protein; 504 AA.  
AC W44156;  
DT 14-MAY-1998 (first entry)

DE Human neuronal nicotinic acetylcholine receptor alpha-3 subunit;  
 KW Human; neuronal nicotinic acetylcholine receptor; alpha-3 subunit;  
 KW brain tissue; screening; NACHR; antibody.  
 OS Homo sapiens.

FT Key Location/Qualifiers  
 FT Peptide 1..30  
 FT Misc\_difference 235 /label= signal  
 FT /label= unspecified  
 FT Domain 240..265 /note= "encoded by WWC"  
 FT /label= TMD1  
 FT /note= "transmembrane domain"  
 FT Misc\_difference 246 /label= unspecified  
 FT /note= "encoded by AWC"  
 FT Domain 273..296 /label= TMD2  
 FT /note= "transmembrane domain"  
 FT Misc\_difference 275 /note= "encoded by TGY"  
 FT Domain 302..326 /label= TMD3  
 FT /note= "transmembrane domain"  
 FT Misc\_difference 309 /label= unspecified  
 FT /note= "encoded by WWC"  
 FT Misc\_difference 347 /label= unspecified  
 FT /note= "encoded by AWC"  
 FT Misc\_difference 354 /label= unspecified  
 FT /note= "encoded by TWC"  
 FT Domain 459..480 /label= TMD4  
 FT /note= "transmembrane domain"  
 FT Region 327..458 /label= cytoplasmic\_loop  
 PN W09420617-A2.  
 PD 15-SEP-1994. U02447.  
 PF 08-MAR-1994; US-028031.  
 PR (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PA (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Elliott KJ, Ellis SB, Harpold MM;  
 DR WPI: 94-303024/37.  
 DR N-PSDB: V12200.

PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
 PT also transformed cells useful for screening cpds. which modulate  
 PT activity of the receptor  
 PS Example 2: Page 72-73: 99pp; English.  
 CC The present sequence represents a human neuronal nicotinic acetylcholine  
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta  
 CC NACHR subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal NACHR.  
 CC Subunit specific antibodies may be used to monitor the distribution  
 CC and expression density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits or specific receptor  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes.  
 SQ Sequence 504 AA;

Query Match 55.2%; Score 32; DB 1; Length 504;  
 Best Local Similarity 58.3%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 YRLAIRRLRY 12

Db 232 YSLXIRRLSLFY 243

| | | | | | | |

RESULT 15  
 W55668  
 ID W55668 standard; Protein; 803 AA.  
 AC W55668;  
 DT 24-JUN-1998 (first entry)  
 DE H. pylori ORF hp461394\_3368767\_c1\_80 cell envelope flagellar protein.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN W09737044-A1.  
 PD 09-OCT-1997.  
 PF 27-MAR-1997; U05223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI: 97-503122/46.  
 DR N-PSDB: V25077.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 14,80; Pages 874-876; 1145pp; English.  
 CC This is the sequence of a H. pylori cell envelope flagellar protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds.  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 SQ Sequence 803 AA;

Query Match 54.3%; Score 31.5; DB 1; Length 803;  
 Best Local Similarity 53.3%; Pred. No. 3e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 1 YRLAIR---RILRY 12  
 | : | | | | |  
 Db 287 YKQAVRYKRILLEY 301

Search completed: February 8, 2000, 01:29:40  
 Job time: 1752 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2000, 11:54:28 ; Search time 117.7 Seconds  
(without alignments)  
4.809 Million cell updates/sec

Title: US-08-653-294-18  
Perfect score: 58  
Sequence: 1 YRLAIRLLRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR62:\*  
1: Pirl:\*  
2: Pirl2:\*  
3: Pirl3:\*  
4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	70.7	348	2 S29990	histocompatibility
2	37	63.8	252	2 I48120	P-glycoprotein - C
3	37	63.8	259	2 I48119	P-glycoprotein - C
4	37	63.8	1281	2 I48123	P-glycoprotein iso
5	36	62.1	141	2 H71504	ribosomal protein
6	36	62.1	309	2 G70882	probable oxidoredu
7	35	60.3	142	2 C72053	L17 ribosomal prot
8	35	60.3	388	2 S15593	hypothetical prote
9	35	60.3	1283	2 A47377	multidrug resist
10	34	58.6	152	2 E72382	iron-sulfur cluste
11	34	58.6	280	2 G69807	conserved hypothet
12	34	58.6	349	2 G70542	probable bios prot
13	34	58.6	690	2 S54211	Armi protein precu
14	34	58.6	735	2 T13646	hypothetical prote
15	34	58.6	814	2 T00953	hypothetical prote
16	34	58.6	1712	2 C71618	hypothetical prote
17	34	58.6	1733	1 PNBV2L	DNA-directed RNA p
18	33	56.9	178	2 F64429	DNA-directed RNA p
19	33	56.9	118	2 S73487	Rnasep C5 chain -
20	33	56.9	133	2 D72110	hypothetical prote
21	33	56.9	168	2 A69422	F420-nonreducing h
22	33	56.9	194	2 T16556	hypothetical prote
23	33	56.9	313	2 JC5342	Na+/H+ antiporter
24	33	56.9	326	2 B71808	type II restrictio
25	33	56.9	334	2 S34254	glyceraldehyde-3-p
26	33	56.9	345	2 S07114	MHC class I histoc
27	33	56.9	362	2 A60384	MHC class I histoc
28	33	56.9	453	2 F71374	probable hemolysin
29	33	56.9	532	2 G70986	probable coA ligas
30	33	56.9	576	2 A71497	probable DNA misma

31	33	56.9	586	2 F64186	ABC-type transport
32	33	56.9	587	2 H64045	probable ABC-type
33	33	56.9	1124	2 D65032	hypothetical prote
34	32	55.2	83	1 W8BPT3	gene 18.7 protein
35	32	55.2	128	2 S69745	hypothetical prote
36	32	55.2	134	2 F72250	ribosomal protein
37	32	55.2	218	2 S07359	regulatory protein
38	32	55.2	218	2 S30287	regulatory protein
39	32	55.2	253	2 G72598	probable ABC trans
40	32	55.2	259	2 A75167	hypothetical prote
41	32	55.2	265	2 S74282	hypothetical prote
42	32	55.2	354	2 E69499	iron-sulfur bindin
43	32	55.2	408	2 D70365	fimbrial assembly
44	32	55.2	421	2 D64441	hypothetical prote
45	32	55.2	448	2 I51883	alpha-2B-adrenergi

ALIGNMENTS

RESULT 1

S29990  
histocompatibility antigen, HLA-F-like - rhesus macaque  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S29990

R:Bontrop, R.R.  
submitted to the EMBL Data Library, February 1993

A:Reference number: S29990  
A:Accession: S29990  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-348 <BON>  
A:Cross-references: EMBL:Z21819; NID:g38568; PIDN:CAA79885.1; PID:g38569  
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology  
C:Keywords: transmembrane protein  
F:219-284/Domain: immunoglobulin homology <IMM>

Query Match 70.7%; Score 41; DB 2; Length 348;  
Best Local Similarity 63.6%; Pred. No. 1.8;  
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RLAIRLLRY 12

Db 98 RVALRKLLRY 108

RESULT 2

I48120  
P-glycoprotein - Chinese hamster (fragment)  
C:Species: Crictetus griseus (Chinese hamster)  
C:Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: I48120  
R:Ng, W.F.; Sarangi, F.; Zastawny, R.L.; Veinot-Drebot, L.; Ling, V.  
Mol. Cell. Biol. 9, 1224-1232, 1989  
A:Title: Identification of members of the P-glycoprotein multigene family.  
A:Reference number: I48119; MUID:89261726  
A:Accession: I48120

A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA

A:Residues: 1-252 <RES>

A:Cross-references: GB:M25792; NID:g576810; PIDN:AAA53440.1; PID:g576813

C:Genetics:

A:Gene: pgp3

A:Introns: 66/3; 135/3; 184/3

C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

C:Keywords: ATP; P-loop

F:25-221/Domain: ATP-binding cassette homology <ABC2>

F:42-49/Region: nucleotide-binding motif A (P-loop)

Query Match 63.8%; Score 37; DB 2; Length 252;

Best Local Similarity 80.0%; Pred. No. 7.8;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11  
Db 155 RLAIIRALLR 164

RESULT 3  
P-glycoprotein - Chinese hamster (fragment)  
C:Species: Cricetulus griseus (Chinese hamster)  
C>Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: I48119  
R:NG, W.F.; Sarangi, F.; Zastawny, R.L.; Veinot-Drebot, L.; Ling, V.  
Mol. Cell. Biol. 9, 1224-1232, 1989  
A:Title: Identification of members of the P-glycoprotein multigene family.  
A:Reference number: I48119; MUID:89261726  
A:Accession: I48119  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-259 <RES>  
A:Cross-references: GB:M25792; NID:g576810; PIDN:AAA53439.1; PID:g576812  
C:Genetics:  
A:Gene: p9p3  
A:Introns: 66/3; 142/3; 191/3  
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology  
C:Keywords: ATP; P-loop  
F:25-228/Domain: ATP-binding cassette homology <ABC2>  
F:42-49/Region: nucleotide-binding motif A (P-loop)

Query Match 63.8%; Score 37; DB 2; Length 259;  
Best Local Similarity 80.0%; Pred. No. 8;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11  
Db 162 RLAIIRALLR 171

RESULT 4  
I48123  
P-glycoprotein isoform III - Chinese hamster  
C:Species: Cricetulus griseus (Chinese hamster)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Aug-1999  
C:Accession: I48123  
R:Endicott, J.A.; Sarangi, F.; Ling, V.  
DNA Seq. 2, 89-101, 1991  
A:Title: Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene family  
A:Reference number: I48121; MUID:92135896  
A:Accession: I48123  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1281 <RES>  
A:Cross-references: GB:M60042; NID:gl91168; PIDN:AAA68885.1; PID:gl91169  
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology  
C:Keywords: ATP; glycoprotein; P-loop  
F:412-606/Domain: ATP-binding cassette homology <ABC1>  
F:429-436/Region: nucleotide-binding motif A (P-loop)  
F:1054-1250/Domain: ATP-binding cassette homology <ABC2>  
F:1071-1078/Region: nucleotide-binding motif A (P-loop)

Query Match 63.8%; Score 37; DB 2; Length 1281;  
Best Local Similarity 80.0%; Pred. No. 37;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11  
Db 1184 RLAIIRALLR 1193

Best Local Similarity 80.0%; Pred. No. 7.8;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11  
Db 155 RLAIIRALLR 164

RESULT 5  
H71504  
ribosomal protein L17 - Chlamydia trachomatis  
C:Species: Chlamydia trachomatis  
C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 13-Aug-1999  
C:Accession: H71504; I40747  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: H71504  
A:Molecule type: DNA  
A:Residues: 1-141 <ARN>  
A:Cross-references: GB:AE001273; GB:AE001273; NID:g3328931; PIDN:AA68107.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
R:Gu, L.; Wenman, W.M.; Remacha, M.; Meuser, R.; Coffin, J.; Kaul, R.  
J. Bacteriol. 177, 2594-2601, 1995  
A:Title: Chlamydia trachomatis RNA polymerase alpha subunit: sequence and structural  
A:Reference number: I40743; MUID:95247702  
A:Accession: I40747  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 20-115, 'R', 117-141 <GUL>  
A:Cross-references: GB:L33834; NID:g620026; PIDN:AAA74990.1; PID:g620030  
C:Genetics:  
A:Gene: rll7  
C:Superfamily: Escherichia coli ribosomal protein L17  
C:Keywords: protein biosynthesis; ribosome

Query Match 62.1%; Score 36; DB 2; Length 141;  
Best Local Similarity 54.5%; Pred. No. 6.9;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 12  
Db 64 RLAVGLMVRV 74

RESULT 6  
G70882  
probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 29-Sep-1999  
C:Accession: G70882  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Nature 393, 537-544, 1998  
A:Authors: Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Qua  
; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: G70882  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-309 <COL>  
A:Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA15591.1; PID:el29  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV2776C  
C:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; fer  
F:9-211/Domain: cytochrome-b5 reductase homology <CBR>  
F:240-297/Domain: ferredoxin [2Fe-2S] homology <FER>

Query Match 62.1%; Score 36; DB 2; Length 309;  
Best Local Similarity 87.5%; Pred. No. 15;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLAIRRI 8  
Db 65 YRLAIRRI 72

```

RESULT 7
C72053
L17 ribosomal protein - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 13-Aug-1999
C:Accession: C72053
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: C72053
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <ARN>
A:Cross-references: GB:AE001647; GB:AE001363; NID:g4376920; PIDN:AAID18764.1; PID:g437692
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: rli7
C:Superfamily: Escherichia coli ribosomal protein L17

Query Match 60.3%; Score 35; DB 2; Length 142;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLRY 12
I:| | | | |
Db 64 RLAIIRILLRY 74

RESULT 8
S15593
hypothetical protein (insertion sequence ISH27-3) - Halobacterium halobium
C:Species: Halobacterium halobium
C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 30-Jun-1998
C:Accession: S15593
R:Pfeifer, F.; Blaselo, U.
Nucleic Acids Res. 18, 6921-6925, 1990
A:Title: Transposition burst of the ISH27 insertion element family in Halobacterium halo
A:Reference number: S15591; MUID:91088266
A:Accession: S15593
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <PFE>
A:Cross-references: EMBL:X54434
A:Note: the authors translated the initiation codon GTG for residue 1 as Val
C:Genetics:
A:Mobile element: insertion sequence ISH27-3
A:Start codon: GTG

Query Match 60.3%; Score 35; DB 2; Length 388;
Best Local Similarity 75.0%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRLAIRRI 8
I:| | | | |
Db 154 YRLAIRRL 161

RESULT 9
A47377
multidrug resistance protein Mdr50 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 19-May-1995 #sequence_revision 19-May-1995 #text_change 10-Jul-1998
C:Accession: A47377
R:Gerrard, B.; Stewart, C.; Dean, M.
Genomics 17, 83-88, 1993
A:Title: Analysis of Mdr50: a Drosophila p-glycoprotein/multidrug resistance gene homolo
A:Reference number: A47377; MUID:94010914
A:Accession: A47377
A>Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-1283 <GER>
A:Cross-references: GB:L07065
C:Genetics:
A:Gene: FlyBase:Mdr49
A:Cross-references: FlyBase:FBgn0010241
A:Introns: 87/3; 182/2; 832/3; 934/3; 981/2; 1168/3; 1217/3
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C:Keywords: ATP; P-loop
F:418-613/Domain: ATP-binding cassette homology <ABC1>
F:435-442/Region: nucleotide-binding motif A (P-loop)
F:1058-1254/Domain: ATP-binding cassette homology <ABC2>
F:1075-1082/Region: nucleotide-binding motif A (P-loop)

Query Match 60.3%; Score 35; DB 2; Length 1283;
Best Local Similarity 70.0%; Pred. No. 91;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11
I:| | | | |
Db 547 RLAIIRALLR 556

RESULT 10
E72382
iron-sulfur cluster-binding protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Sep-1999
C:Accession: E72382
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: E72382
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <ARN>
A:Cross-references: GB:AE001719; GB:AE000512; NID:g4980893; PID:g4980900; TIGR:TM0396
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0396
C:Superfamily: nrfC protein; ferredoxin 2[4Fe-4S] homology

Query Match 58.6%; Score 34; DB 2; Length 152;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 LAIRILLR 11
I:| | | | |
Db 3 LAIRILLR 11

RESULT 11
G69807
conserved hypothetical protein yfkC - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C:Accession: G69807
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

```

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: G69807  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-280 <KUN>  
 A:Cross-references: GB:299108; GB:AL009126; NID:g2633055; PID:el182784; PID:g2633118  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yfkC

Query Match 58.6% Score 34; DB 2; Length 280;  
 Best Local Similarity 66.7% Pred. No. 32;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 LAIRRIILLR 11  
 |||:|:|:  
 Db 252 LAVRVLVR 260

RESULT 12  
 G70342

probable bioB protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 16-Jul-1999  
 C:Accession: G70342  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70300; MUID:98295987  
 A:Accession: G70342  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-349 <COL>  
 A:Cross-references: GB:295586; GB:AL123456; NID:g3261785; PIDN:CAB09080.1; PID:g3171168;  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: bioB  
 C:Superfamily: biotin synthetase

Query Match 58.6% Score 34; DB 2; Length 349;  
 Best Local Similarity 50.0% Pred. No. 40;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRLAIRILLRY 12  
 |||:|:|:  
 Db 280 FRLALPTMLRF 291

RESULT 13  
 S54211

ATM1 protein precursor - yeast (*Saccharomyces cerevisiae*)  
 A:Alternate names: MDY protein; protein YW9552.03c; protein YMR301c  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999  
 C:Accession: S54211; S53971; S51801  
 R:Kispal, G.; Lill, R.; Neupert, N.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S54211  
 A:Accession: S54211  
 A:Molecule type: DNA  
 A:Residues: 1-690 <KIS>  
 A:Cross-references: EMBL:X81715; NID:g793876; PIDN:CAA57359.1; PID:g793877  
 R:Connor, R.; Churcher, C.M.  
 submitted to the EMBL Data Library, April 1995

A:Reference number: S53969  
 A:Accession: S53971  
 A:Molecule type: DNA  
 A:Residues: 1-690 <CON>  
 A:Cross-references: EMBL:Z49212; NID:g798940; PIDN:CAA89134.1; PID:g798943; MIPS:YMR3  
 R:Leighton, J.; Schatz, G.  
 EMBO J. 14, 188-195, 1995  
 A:Title: An ABC transporter in the mitochondrial inner membrane is required for norma  
 A:Reference number: S51801; MUID:95129546  
 A:Accession: S51801  
 A:Molecule type: DNA  
 A:Residues: 1-27, 'RNHS', 28-690 <LEI>  
 A:Cross-references: EMBL:X82612; NID:g575392; PIDN:CAA57938.1; PID:g575393  
 C:Genetics:  
 A:Gene: SGD:ATM1; MDY  
 A:Cross-references: SGD:S0004916; MIPS:YMR301c  
 A:Map position: 13R  
 A:Genome: nuclear  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog  
 C:Keywords: ATP; mitochondrial inner membrane; mitochondrion; P-loop; transmembrane p  
 F:111-127/Domain: transmembrane #status predicted <TM1>  
 F:148-164/Domain: transmembrane #status predicted <TM2>  
 F:233-249/Domain: transmembrane #status predicted <TM3>  
 F:258-274/Domain: transmembrane #status predicted <TM4>  
 F:452-648/Domain: ATP-binding cassette homolog <ABC>  
 F:469-476/Region: nucleotide-binding motif A (P-loop)

Query Match 58.6% Score 34; DB 2; Length 690;  
 Best Local Similarity 70.0% Pred. No. 77;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11  
 |||:|:|:  
 Db 580 RLAIARVLK 589

RESULT 14

T13646  
 hypothetical protein EG:95B7.9 - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999  
 C:Accession: T13646  
 R:Ferraz, C.; Vidal, S.; Brun, C.; Bucheton, A.; Demaille, J.G.  
 submitted to the EMBL Data Library, April 1999  
 A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.  
 A:Reference number: Z17694  
 A:Accession: T13646  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-735 <FER>  
 A:Cross-references: EMBL:AL021728; NID:el355643; PID:el301392; PIDN:CAA16821.1  
 C:Genetics:  
 A:Introns: 39/3; 93/2; 294/2; 387/2  
 A:Note: EG:95B7.9

Query Match 58.6% Score 34; DB 2; Length 735;  
 Best Local Similarity 75.0% Pred. No. 82;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 IRRILLRY 12  
 |||:|:|:  
 Db 473 IRRILLRY 480

RESULT 15

T00953  
 hypothetical protein F20D22.3 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cross)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Sep-1999  
 C:Accession: T00953  
 R:Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji



K.; Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.;  
submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence.  
A:Reference number: Z14214  
A:Accession: T00953  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-814 <YYS>  
A:Cross-references: EMBL:AC002411; NID:g2570223; PID:g3142290; GSPDB:GN0059; ATSP:F20D22  
C:Genetics:  
A:Gene: ATSP:F20D22.3  
A:Map position: 1  
A:Introns: 64/1; 144/1; 239/3; 304/1; 386/1; 415/2; 473/3; 516/3; 554/3; 594/2; 624/3; 6

Query Match 58.6%; Score 34; DB 2; Length 814;  
Best Local Similarity 60.0%; Pred. No. 91;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 LAIRRLRY 12  
| :||| |  
DB 263 LVVRLRLNY 272

Search completed: February 7, 2000, 11:54:29  
Job time: 24339 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 00:59:56 ; Search time 63.71 Seconds  
(without alignments)  
5.625 Million cell updates/sec

Title: US-08-653-294-18  
Perfect score: 58  
Sequence: 1 YRLAIRILLRY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	70.7	348	1 HLA_F_MACMU	P33617 macaca mula
2	37	63.8	1281	1 MDR3_CRIGR	P23174 cricetus
3	36	62.1	122	1 RL17_CHLTR	P47760 chlamydia t
4	36	62.1	621	1 CRT1_CERN	P48537 cercospora
5	34	58.6	349	1 BI0B_MYCTU	O06601 mycobacteri
6	34	58.6	690	1 AT01_YEAST	P40416 saccharomyc
7	34	58.6	1733	1 RPB1_YEAST	P04050 saccharomyc
8	33	56.9	78	1 RPOH_METJA	Q58443 methanococ
9	33	56.9	118	1 RNPA_MYCPN	P75111 mycoplasma
10	33	56.9	334	1 G3P_CLOPA	Q59309 clostridium
11	33	56.9	346	1 IC28_PANTR	P16215 pan troglod
12	33	56.9	362	1 HLA_F_HUMAN	P30511 homo sapien
13	33	56.9	586	1 CYDD_HAEN	P45082 haemophilus
14	33	56.9	587	1 MSBA_HAEN	P44407 haemophilus
15	33	56.9	1124	1 YPHG_ECOLI	P76585 escherichia
16	32	55.2	83	1 V187_BPT3	P10302 bacterioph
17	32	55.2	212	1 YCD4_YEAST	P25369 saccharomyc
18	32	55.2	217	1 TER4_ECOLI	P09164 escherichia
19	32	55.2	218	1 TER8_PASPI	P51562 pasteurella
20	32	55.2	411	1 VHS5_CAEEL	O18304 caenorhabdi
21	32	55.2	448	1 AZAB_CAVPO	Q60475 cavia porce
22	32	55.2	450	1 AZAB_HUMAN	P18089 homo sapien
23	32	55.2	453	1 AZAB_RAT	P19328 rattus norv
24	32	55.2	455	1 AZAB_MOUSE	P30545 mus musculi
25	32	55.2	469	1 Y221_MYCTU	P96403 mycobacteri
26	32	55.2	477	1 BGL2_BACSU	P42403 bacillus su
27	32	55.2	508	1 PGKD_TRYBB	P08892 trypanosoma
28	32	55.2	693	1 YETI_SCHPO	O14286 schizosacch
29	31	53.4	128	1 RNPA_MYCCE	P47703 mycoplasma
30	31	53.4	185	1 DPL_MOUSE	Q60870 mus musculi
31	31	53.4	185	1 RPO6_VACCV	P07391 vaccinia vi
32	31	53.4	185	1 RPO6_VARY	P33054 variola vir
33	31	53.4	203	1 NH10_YEAST	Q03435 saccharomyc
34	31	53.4	306	1 BCRA_BACLI	P42332 bacillus li

35	31	53.4	306	1 OPPB_ECOLI	P1132 escherichia
36	31	53.4	344	1 VGLM_HSV6U	Q04630 herpes simp
37	31	53.4	344	1 VGLM_HSV6Z	P52449 herpes simp
38	31	53.4	354	1 Y101_NPVOP	P24653 orgyia pseu
39	31	53.4	375	1 HRMA_PSESY	Q08370 pseudomonas
40	31	53.4	407	1 YAEI_SCHPO	Q09842 schizosacch
41	31	53.4	532	1 HEPA_ANASP	P22638 anabaena sp
42	31	53.4	582	1 MSBA_ECOLI	P27299 escherichia
43	31	53.4	628	1 SVI_NOSLO	Q27707 nosena locu
44	31	53.4	681	1 SSAV_SALTY	P74856 salmonella
45	31	53.4	695	1 PRC_HAEN	P45306 haemophilus

#### ALIGNMENTS

RESULT 1  
HLAF\_MACMU STANDARD; PRT; 348 AA.  
AC P33617;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F  
DE ANTIGEN) (LEUKOCYTE ANTIGEN F).  
GN HLA-F OR HLA-F.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopithecinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93246295.  
RA OTTING N., BONTROP R.E.;  
RT "Characterization of the rhesus macaque (Macaca mulatta) equivalent  
RT of HLA-F";  
RL Immunogenetics 38:141-145(1993).  
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 221819; CAA79885.1; -.  
CC PIR; S29990; S29990.  
CC HSSP; P03989; LHSA.  
CC PROSITE; PS00290; IG\_MHC; 1.  
CC PFAM; PF00047; 19; 1.  
CC PFAM; PF00129; MHC-I; 1.  
CC MHC I; Transmembrane; Glycoprotein; Signal.  
CC SIGNAL 1 21  
CC CHAIN 22 348  
CC BY SIMILARITY.  
CC HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
CC ALPHA CHAIN F.  
CC DOMAIN 22 113  
CC DOMAIN 114 205  
CC DOMAIN 206 297  
CC DOMAIN 298 307  
CC TRANSMEM 308 331  
CC DOMAIN 332 348  
CC DISULFID 124 187  
CC DISULFID 226 282  
CC CARBOHYD 109 109  
CC POTENTIAL.  
CC SEQUENCE 348 AA; 39300 MW; 3A375142 CRC32;

Query Match

70.7% ; Score 41; DB 1; Length 348;

```

Best Local Similarity 63.8%; Pred. No. 0.52;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLAIIRILLRY 12
   |||:|:|:|:|:|
Db 98 RVALRKLLRY 108

RESULT 2
MDR3_CRIGR STANDARD; PRT; 1281 AA.
AC P23174;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPROTEIN 3).
GN PGY3 OR PGP3
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RX MEDLINE: 92135896.
RA ENDICOTT J.A., SARANGI F., LING V.;
RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
   gene family.";
RL DNA Seq. 2:89-101(1991).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CC CANNOT.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M60042; AAA68885.1; -
CC HSSP: P13569; INBD.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC PFAM: PF00005; ABC_tran; 2.
CC PFAM: PF00864; ABC_membrane; 2.
CC ATP-binding: Glycoprotein; Transmembrane; Transport; Duplication;
CC Multigene family.
CC DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 58 78 POTENTIAL.
CC FT TRANSMEM 122 142 POTENTIAL.
CC FT TRANSMEM 191 211 POTENTIAL.
CC FT TRANSMEM 218 238 POTENTIAL.
CC FT TRANSMEM 299 319 POTENTIAL.
CC FT TRANSMEM 328 348 POTENTIAL.
CC FT TRANSMEM 349 712 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 713 733 POTENTIAL.
CC FT TRANSMEM 733 733 POTENTIAL.
CC FT TRANSMEM 758 778 POTENTIAL.
CC FT TRANSMEM 834 854 POTENTIAL.
CC FT TRANSMEM 855 875 POTENTIAL.
CC FT TRANSMEM 938 958 POTENTIAL.
CC FT TRANSMEM 975 995 POTENTIAL.
CC FT TRANSMEM 995 1281 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 1281 1078 ATP (POTENTIAL).
CC FT NP_BIND 429 436 ATP (POTENTIAL).
CC FT NP_BIND 1071 1078 ATP (POTENTIAL).
CC SEQUENCE 1281 AA; 140866 MW; 9FE9DF5C CRC32;

Query Match 63.8%; Score 37; DB 1; Length 1281;
Best Local Similarity 80.0%; Pred. No. 14;

Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLAIIRILLRY 12
   |||:|:|:|:|:|
Db 98 RVALRKLLRY 108

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11
   |||:|:|:|:|:|
Db 1184 RLAIIRALLR 1193

RESULT 3
RL17_CHLTR STANDARD; PRT; 122 AA.
AC P47760;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L17.
GN RPLO.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RX MEDLINE: 95247702.
RA GU L.J., WENMAN W.M., REMACHA M., MEUSER R.U., COFFIN J.M., KAUL R.;
RT "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and
   structural analysis.";
RL J. Bacteriol. 177:2594-2601(1995).
CC -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L33834; AAA74990.1; -
CC PROSITE: PS01167; RIBOSOMAL_L17; 1.
CC PFAM: PF01196; Ribosomal_L17; 1.
CC Ribosomal protein.
CC SEQUENCE 122 AA; 13969 MW; B8C43F7D CRC32;

Query Match 62.1%; Score 36; DB 1; Length 122;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLRY 12
   |||:|:|:|:|:|
Db 45 RLAVGRLLVRY 55

RESULT 4
CRTI_CERNC STANDARD; PRT; 621 AA.
AC P48537;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PHYTOENE DEHYDROGENASE (EC 1.3.-.-) (PHYTOENE DESATURASE).
GN PDH1.
OS Cercospora nicotianae.
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;
OC Pleosporales; Leptosphaeriaceae; mitosporic Leptosphaeriaceae;
OC Cercospora.
RN [1]
RX MEDLINE: 94368091.
RA EHRENSHAFT M., DAUB M.E.;
RT "Isolation, sequence, and characterization of the Cercospora
   nicotianae phytoene dehydrogenase gene.";
RL Appl. Environ. Microbiol. 60:2766-2771(1994).

```

CC -!- FUNCTION: THIS ENZYME CONVERTS PHYTOENE INTO ZETA-CAROTENE VIA THE  
CC INTERMEDIARY OF PHYTOFLUENE BY THE SYMMETRICAL INTRODUCTION OF TWO  
CC DOUBLE BONDS AT THE C-11 AND C-11' POSITIONS OF PHYTOENE.  
CC -!- COFACTOR: FAD (PROBABLE).  
CC -!- PATHWAY: CAROTENOID BIOSYNTHESIS.  
CC -!- SIMILARITY: TO BACTERIAL PHYTOENE DEHYDROGENASES AND TO BACTERIAL  
CC METHOXYNEUROSPORENE DEHYDROGENASE (CRTD).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U03903; AAB86988.1; -.  
CC PROSITE; PS00982; PHYTOENE\_DH: 1.  
CC Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;  
CC Transmembrane.  
CC NP\_BIND 11 44 FAD (ADP PART) (POTENTIAL).  
CC TRANSMEM 536 556 POTENTIAL.  
CC SEQUENCE 621 AA; 69529 MW; CDA97504 CRC32;  
CC  
CC Query Match 62.1%; Score 36; DB 1; Length 621;  
CC Best Local Similarity 54.5%; Pred. No. 10;  
CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
CC  
CC Qy 1 YRLAIRILLR 11  
CC | : : : : :  
CC Db 140 YELSVRELLR 150  
CC  
CC RESULT 5  
CC BIOB\_MYCTU STANDARD; PRT; 349 AA.  
CC AC 008601:  
CC DT 15-JUL-1998 (Rel. 36, Created)  
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)  
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
CC DE BIOTIN SYNTHASE (EC 2.8.1.6) (BIOTIN SYNTHETASE).  
CC GN BIOB OR RV1589 OR WICY336.15C.  
CC OS Mycobacterium tuberculosis.  
CC OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
CC OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
CC [1]  
CC RN SEQUENCE FROM N.A.  
CC RC STRAIN-H37RV.  
CC RX MEDLINE; 98295987.  
CC RA COLE S.T., BROSH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,  
CC RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,  
CC RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,  
CC RA DAVIES R., DEVLIN K., FELTWEILL T., GENTLES S., HAMLIN N., HOLROYD S.,  
CC RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,  
CC RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,  
CC RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,  
CC RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;  
CC RT Deciphering the biology of Mycobacterium tuberculosis from the  
CC RT complete genome sequence."  
CC RL Nature 393:537-544(1998).  
CC [2]  
CC RN SEQUENCE FROM N.A.  
CC RC STRAIN-PASTEUR;  
CC RA YU S., JACOBS W.R. JR.;  
CC RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
CC CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) - BIOTIN.  
CC CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.  
CC CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z95586; CAB09080.1; -.  
CC DR EMBL; AF041819; AAB96962.1; -.  
CC KW Biotin biosynthesis; Iron-sulfur; Transferase.  
CC FT METAL 85 85 IRON-SULFUR (POTENTIAL).  
CC FT METAL 89 89 IRON-SULFUR (POTENTIAL).  
CC FT METAL 92 92 IRON-SULFUR (POTENTIAL).  
CC SQ SEQUENCE 349 AA; 37550 MW; 31B0B9DA CRC32;  
CC  
CC Query Match 58.6%; Score 34; DB 1; Length 349;  
CC Best Local Similarity 50.0%; Pred. No. 14;  
CC Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
CC  
CC Qy 1 YRLAIRILLRY 12  
CC : : : : :  
CC Db 280 FRALPRTMLRF 291  
CC  
CC RESULT 6  
CC ATMI\_YEAST STANDARD; PRT; 690 AA.  
CC ID ATMI\_YEAST  
CC AC P40416;  
CC DT 01-FEB-1995 (Rel. 31, Created)  
CC DT 01-FEB-1996 (Rel. 33, Last sequence update)  
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
CC DE MITOCHONDRIAL TRANSPORTER ATMI PRECURSOR.  
CC GN ATMI OR MDY OR YMR301C OR YMR952.03C.  
CC OS Saccharomyces cerevisiae (Baker's yeast).  
CC OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
CC OC Saccharomycetaceae; Saccharomycetes.  
CC [1]  
CC RN SEQUENCE FROM N.A.  
CC RC STRAIN-JK9-3D ALPHA;  
CC RX MEDLINE; 95129546.  
CC RA LEIGHTON J., SCHATZ G.;  
CC RT "An ABC transporter in the mitochondrial inner membrane is required  
CC RT for normal growth of yeast."  
CC RL EMOB J. 14:188-195(1995).  
CC [2]  
CC RN SEQUENCE FROM N.A.  
CC RX MEDLINE; 98089018.  
CC RA KISPAL G., CSERE P., GUIARD B., LILL R.;  
CC RT "The ABC transporter Acm1p is required for mitochondrial iron  
CC RT homeostasis."  
CC RL FEBS Lett. 418:346-350(1997).  
CC [3]  
CC RN SEQUENCE FROM N.A.  
CC RC STRAIN-S288C / AB972;  
CC RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
CC RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
CC CC -!- FUNCTION: PROBABLE TRANSPORTER FOR A YET UNKNOWN SUBSTRATE.  
CC REQUIRED FOR MITOCHONDRIAL IRON HOMEOSTASIS.  
CC CC -!- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).  
CC CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE.  
CC CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS). MDR SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X82612; CAA57938.1; -.  
CC

```

DR EMBL; X81715; CAA57359.1; -.
DR EMBL; 249212; CAA89134.1; -.
DR SGD; L0000140; ATMI.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1.
DR PFAM; PF00684; ABC_membrane; 1.
DR ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane;
KW Transmembrane; Transport; Mitochondrion; Inner membrane;
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN 1 ? 690 MITOCHONDRIAL TRANSPORTER ATMI.
FT TRANSEM 112 132 POTENTIAL.
FT TRANSEM 152 172 POTENTIAL.
FT TRANSEM 229 249 POTENTIAL.
FT TRANSEM 254 274 POTENTIAL.
FT TRANSEM 348 368 POTENTIAL.
FT TRANSEM 375 395 POTENTIAL.
FT NP_BIND 469 476 ATP (POTENTIAL).
FT CONFLICT 27 27 S -> SRNHS (IN REF. 1).
SQ SEQUENCE 690 AA; 77522 MW; 2E8AEC70 CRC32;

Query Match 58.6%; Score 34; DB 1; Length 690;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11
DB 580 RLAIARVLK 589

RESULT 7
RPBL_YEAST STANDARD; PRT; 1733 AA.
AC P04050; Q12364; Q92315;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (B220).
GN RPBI OR RPO21 OR RPB220 OR SUA8 OR YDL140C OR D2150.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A364A;
RX MEDLINE; 85282617.
RA ALLISON L.A., MOYLE M., SHALES M., INGLES C.J.;
RT "Extensive homology among the largest subunits of eukaryotic and
RL Cell 42:599-610(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE; 97127826.
RA WOELFL S., HANEMAN V., SALUZ H.P.;
RT "Analysis of a 26,756 bp segment from the left arm of yeast
RL Chromosome IV.";
RN Yeast 12:1549-1554(1996).
RP [3]
RP SEQUENCE OF 1669-1733 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE; 95377607.
RA CRONAN J.E. JR., WALLACE J.C.;
RT "The gene encoding the biotin-apolipoprotein ligase of Saccharomyces
RL cerevisiae.";
RC FEMS Microbiol. Lett. 130:221-230(1995).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
RNA(N).
CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.

THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
-!- SUBCELLULAR LOCATION: NUCLEAR.
-!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
-!- THE PHOSPHORYLATION ACTIVATES POL2.
-!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
III FOR 5S AND TRNA GENES.
-!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X03128; CAA26904.1; -.
CC EMBL; X96876; CAA65619.1; -.
CC EMBL; 274188; CAA98713.1; -.
CC EMBL; U27182; AAC49058.1; -.
CC PIR; A00692; RNBX2L.
CC SGD; L0001744; RPO21.
CC PROSITE; PS00115; RNA_POL_II_REPEAT; 22.
CC PFAM; PF00623; RNA_POL_A; 1.
CC Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 67 83 C2H2-TYPE (POTENTIAL).
FT DOMAIN 1544 1719 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT VARIANT 1653 1659 MISSING (IN STRAIN A364A).
FT CONFLICT 1514 1514 A -> V (IN REF. 1).
FT CONFLICT 1524 1524 G -> A (IN REF. 1).
FT CONFLICT 1601 1601 T -> M (IN REF. 1).
SQ SEQUENCE 1733 AA; 191610 MW; BB65D7EE CRC32;

Query Match 58.6%; Score 34; DB 1; Length 1733;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 12
DB 1025 RLAIARVLQ 1035

RESULT 8
RPOH_METJA STANDARD; PRT; 78 AA.
AC Q58443;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT H (EC 2.7.7.6).
GN RPOH OR M1039.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEORGAGAN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., RAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RN Science 273:1058-1073(1996).

```

CC [2] STRUCTURE BY NMR.  
RX MEDLINE: 93208760.  
RA THIRU A., HODACH M., ELORANTA J.J., KOSTOUROU V., WEINZIERL R.O.,  
RT MATTHEWS S.;  
RT "RNA polymerase subunit H features a beta-ribbon motif within a novel  
fold that is present in archaea and eukaryotes.";  
RL J. Mol. Biol. 287:753-760(1999).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE + N PYROPHOSPHATE +  
RNA(N).  
CC -!- SIMILARITY: TO OTHER ARCHAEABACTERIAL RPOH AND TO THE C-TERMINAL  
OF EUKARYOTIC SUBUNIT ABC27 (RPS5).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: U67546; AAB99042.1; -.  
DR PDB: 1HMJ; 05-APR-99.  
DR TIGR: M31039; -.  
DR PROSITE: PS01110; RNA\_POL\_H\_23KD; 1.  
DR PFAM: PF01191; RNA\_pol\_H\_1.  
KW Transferase; Transcription; DNA-directed RNA polymerase; 3D-structure.  
SQ SEQUENCE 78 AA; 9001 MW; 9F10C0F3 CRC32;  
  
Query Match 56.9%; Score 33; DB 1; Length 78;  
Best Local Similarity 66.7%; Pred. No. 4.2;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 YRLAIRRL 9  
DB 70 YRLVIRRII 78  
  
RESULT 9  
RNPA\_MYCPN STANDARD; PRT; 118 AA.  
AC P75111;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE RIBONUCLEASE P PROTEIN COMPONENT (EC 3.1.26.5) (PROTEIN C5) (RNASE P).  
GN RNPA.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
CC [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE: 97105885.  
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,  
RA HERRMANN R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -!- FUNCTION: RIBONUCLEASE P GENERATES MATURE TRNA MOLECULES BY  
CLEAVING THEIR 5' ENDS. IT CAN CLEAVE ALSO THE 4.5S RNA  
(BY SIMILARITY)  
CC -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE OF RNA, REMOVING  
5'-EXTRA-NUCLEOTIDE FROM TRNA PRECURSOR.  
CC -!- MISCELLANEOUS: RNASE P CONSISTS OF A RNA MOIETY (M1, RNPB) AND THE  
PROTEIN COMPONENT. BOTH ARE NECESSARY FOR FULL ENZYMIC ACTIVITY.  
CC HOWEVER, IT IS THE RNA THAT CARRIES THE CATALYTIC SITE (BY  
SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE RNPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: AB000017; AAB95809.1; -.  
DR PROSITE: PS00648; RIBONUCLEASE\_P; 1.  
DR PFAM: PF00825; Ribonuclease\_P; 1.  
KW Hydrolase; Nuclease; Endonuclease; tRNA processing.  
SQ SEQUENCE 118 AA; 14003 MW; 5DB7B1B1 CRC32;  
  
Query Match 56.9%; Score 33; DB 1; Length 118;  
Best Local Similarity 54.5%; Pred. No. 6.6;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 YRLAIRRIIR 11  
DB 53 YKLAVERNLR 63  
  
RESULT 10  
G3P\_CLOPA STANDARD; PRT; 334 AA.  
AC Q59309;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH) (CP  
17/CP 18).  
GN GAP.  
OS Clostridium pasteurianum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
CC [1]  
RP SEQUENCE FROM N.A.  
RA OSTER T., ASSOBEI O., SCHERRER S., BRANLANT G., BRANLANT C.;  
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP PARTIAL SEQUENCE OF 1-26.  
RC STRAIN=W5;  
RX MEDLINE: 98291870.  
RA FLENGSRUD R., SKJELDAL L.;  
RT "Two-dimensional gel electrophoresis separation and N-terminal  
sequence analysis of proteins from Clostridium pasteurianum W5.";  
RL Electrophoresis 19:802-806(1998).  
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
+ NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
DEHYDROGENASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: X72219; CAAS1020.1; -.  
DR HSP: P00362; 2GDI.  
DR PROSITE: PS00071; GAPDH; 1.  
DR PFAM: PF00044; gpdh; 1.  
KW Glycolysis; Oxidoreductase; NAD.  
FT BINDING 150 150 GLYCERALDEHYDE 3-PHOSPHATE.  
FT ACT\_SITE 177 177 ACTIVATES THIOL GROUP DURING CATALYSIS.

SQ SEQUENCE 334 AA; 36078 MW; 9061EAF0 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 334;  
Best Local Similarity 87.5%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLAIIRIL 9  
|:|:|:|:|  
DB 14 RLALRIL 21

RESULT 11  
1C28\_PANTR STANDARD; PRT; 346 AA.  
AC P16215;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CH28 ALPHA CHAIN PRECURSOR.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Pan.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90201944.  
RA LAWLOR D.A., WARREN E., WARD F.E., PARHAM P.;  
RT "Comparison of class I MHC alleles in humans and apes.";  
RL Immunol. Rev. 113:147-185(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88319000.  
RA LAWLOR D.A., WARD F.E., ENNIS P.D., JACKSON A.P., PARHAM P.;  
RT "HLA-A and B polymorphisms predate the divergence of humans and  
chimpanzees";  
RL Nature 335:268-271(1988).

CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
MICROGLOBULIN).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M30685; AAA87973.1; -  
DR HSSP; P03989; IHSA.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PFAM; PF00047; Ig; 1.  
DR PFAM; PF00129; MHC.I; 1.  
KW MHC I; Transmembrane; Glycoprotein; Signal.

FT SIGNAL 1 21  
FT CHAIN 22 346  
FT FT CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
CH28 ALPHA CHAIN.  
FT DOMAIN 22 111  
FT DOMAIN 112 203  
FT DOMAIN 204 295  
FT DOMAIN 296 305  
FT DOMAIN 306 329  
FT TRANSMEM 306 329  
FT DOMAIN 330 346  
FT DOMAIN 330 346  
FT DISULFID 122 185  
FT DISULFID 224 280  
FT DISULFID 107 107  
FT CARBOHYD 107 107  
SQ SEQUENCE 346 AA; 39084 MW; 78B65140 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 346;  
Best Local Similarity 54.5%; Pred. No. 21;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 RLAIIRILLRY 12  
|:|:|:|:|  
DB 96 RVALRNLRLRY 106

RESULT 12  
HLAF\_HUMAN STANDARD; PRT; 362 AA.  
AC P30511;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F  
ANTIGEN) (LEUKOCYTE ANTIGEN F) (CDA12).  
GN HLA-F OR HLAF OR HLA-5.4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90111605.  
RA GERAGHTY D.E., WEI X., ORR H.T., KOLLER B.H.;  
RT "Human leukocyte antigen F (HLA-F). An expressed HLA gene composed of  
a class I coding sequence linked to a novel transcribed repetitive  
element.";  
RL J. Exp. Med. 171:1-18(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91197889.  
RA LORY D., EPSTEIN H., HOLMES N.;  
RT "The human class I MHC gene HLA-F is expressed in lymphocytes.";  
RL Int. Immunol. 2:531-537(1990).  
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
MICROGLOBULIN).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X17093; CAA34947.1; -  
DR PIR; A60384; A60384.  
DR PIR; JLO147; JLO147.  
DR HSSP; P03989; IHSA.  
DR MIM; 143110; -  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PFAM; PF00047; Ig; 1.  
DR PFAM; PF00129; MHC.I; 1.  
KW MHC I; Transmembrane; Glycoprotein; Signal.

FT SIGNAL 1 21  
FT CHAIN 22 362  
FT FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
ALPHA CHAIN F.  
FT DOMAIN 22 111  
FT DOMAIN 112 203  
FT DOMAIN 204 295  
FT DOMAIN 296 305  
FT DOMAIN 306 329  
FT TRANSMEM 306 329  
FT DOMAIN 330 362  
FT DOMAIN 330 362  
FT DISULFID 122 185  
FT DISULFID 224 280  
FT DISULFID 107 107  
FT CARBOHYD 107 107  
SQ SEQUENCE 362 AA; 40568 MW; E9B29521 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 362;  
Best Local Similarity 54.5%; Pred. No. 23;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;



Oy 2 RLAIIRLLRY 12  
|:|:|:|:|  
Db 96 RVALRNLRLRY 106

## RESULT 13

CYDD\_HAEIN STANDARD; PRT; 586 AA.  
AC P45082;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE TRANSPORT ATP-BINDING PROTEIN CYDD.  
GN CYDD OR H1157.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.

[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE; 95350630.  
RA FLEISCHMANN R.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,  
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,  
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).

CC -1- FUNCTION: SOMEHOW INVOLVED IN THE CYTOCHROME D BRANCH OF AEROBIC  
CC RESPIRATION. SEEMS TO BE A COMPONENT OF A TRANSPORT SYSTEM  
CC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS). MDR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U32795; AAC22812.1; -.  
DR HSSP; P13569; 1NBD.  
DR TIGR; H1157; -.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR PFAM; PF00005; ABC\_tran; 1.  
DR PFAM; PF00664; ABC\_membrane; 1.  
KW ATP-binding; Transport; Transmembrane; Inner membrane.  
FT TRANSMEM 28 48 POTENTIAL.  
FT TRANSMEM 62 82 POTENTIAL.  
FT TRANSMEM 146 166 POTENTIAL.  
FT TRANSMEM 167 187 POTENTIAL.  
FT TRANSMEM 250 270 POTENTIAL.  
FT TRANSMEM 278 298 POTENTIAL.  
FT NP\_BIND 383 390 ATP (POTENTIAL).  
SQ SEQUENCE 586 AA; 65645 MW; 80EB2DE1 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 586;  
Best Local Similarity 80.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 RLAIIRLLRY 11  
|:|:|:|:|

Db 491 RLAIARALLR 500

## RESULT 14

MSBA\_HAEIN STANDARD; PRT; 587 AA.  
AC P44407;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROBABLE TRANSPORT ATP-BINDING PROTEIN MSBA.  
GN MSBA OR MSH-1 OR H10060.  
OS Haemophilus influenzae;  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.

[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE; 95350630.  
RA FLEISCHMANN R.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,  
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,  
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).

[2]  
RP SEQUENCE OF 1-458 FROM N.A.  
RC STRAIN-BC200;  
RX MEDLINE; 94341577.

RA CLIFTON S.W., MCCARTHY D., ROE B.A.;  
RT "Sequence of the rec-2 locus of Haemophilus influenzae: homologies to  
RT come-ORF3 of Bacillus subtilis and msba of Escherichia coli.";  
RL Gene 146:95-100(1994).  
CC -1- FUNCTION: PROBABLY INVOLVED IN THE SECRETION OF LIPID A (BY  
CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS). MSBA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U32691; AAC21738.1; -.  
DR EMBL; L20805; AAC13734.1; -.  
DR TIGR; H10060; -.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR PFAM; PF00005; ABC\_tran; 1.  
DR PFAM; PF00664; ABC\_membrane; 1.  
KW ATP-binding; Transport; Inner membrane; Transmembrane.  
FT TRANSMEM 20 40 POTENTIAL.  
FT TRANSMEM 68 88 POTENTIAL.  
FT TRANSMEM 157 177 POTENTIAL.  
FT TRANSMEM 257 277 POTENTIAL.  
FT TRANSMEM 279 299 POTENTIAL.  
FT NP\_BIND 381 388 ATP (POTENTIAL).  
SQ SEQUENCE 587 AA; 64912 MW; FC711A0B CRC32;

Query Match 56.9%; Score 33; DB 1; Length 587;  
Best Local Similarity 80.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLAIIRILLR 11  
 |||||  
 Db 493 RLAIARALLR 502

## RESULT 15

YPHG\_ECOLI  
 ID YPHG\_ECOLI STANDARD; PRT; 1124 AA.  
 AC P76585;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 127.3 KD PROTEIN IN CSIE-GLYA INTERGENIC REGION.  
 GN YPHG.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1685;  
 RX MEDLINE: 97426617.  
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., ROSE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE000341; AAC75602.1; -  
 DR ECOGENE; EGI3468; yphg.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1124 AA; 127284 MW; 0A06B4C6 CRC32;

Query Match 56.9%; Score 33; DB 1; Length 1124;  
 Best Local Similarity 87.5%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRLAIREI 8  
 |||||  
 Db 412 YRLAIREI 419

Search completed: February 8, 2000, 00:59:57  
 Job time: 3786 sec